

AN INFORMATION THEORY APPROACH TO WIRELESS SENSOR NETWORK DESIGN

A Thesis
Presented to
The Academic Faculty

by

Bryan Larish

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Electrical and Computer Engineering

Georgia Institute of Technology
May 2013

AN INFORMATION THEORY APPROACH TO WIRELESS SENSOR NETWORK DESIGN

Approved by:

Professor George Riley, Advisor
School of Electrical and Computer
Engineering
Georgia Institute of Technology

Professor Steven McLaughlin
School of Electrical and Computer
Engineering
Georgia Institute of Technology

Professor Yorai Wardi
School of Electrical and Computer
Engineering
Georgia Institute of Technology

Professor James McClellan
School of Electrical and Computer
Engineering
Georgia Institute of Technology

Professor Kishore Ramachandran
School of Computer Science
Georgia Institute of Technology

Dr. William Rix
Office of the Chief Engineer
*Space and Naval Warfare Systems
Command*

Date Approved: 10 December 2012

ACKNOWLEDGEMENTS

I want to thank my research advisor, committee, and family for their help and support.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	iii
LIST OF FIGURES	vi
SUMMARY	viii
I INTRODUCTION	1
1.1 Background	1
1.2 WSN Model	5
1.3 WSN Coding and the Collection Channel	9
1.3.1 Collection Channel	10
1.3.2 WSN Coding	11
1.4 Positive Coding Theorems vs. Negative Coding Theorems	12
II RELATED WORK	13
2.1 Multi-Terminal Coding	15
2.2 Detection	18
2.2.1 WSN Coverage	18
2.2.2 Distributed Detection	18
2.3 Estimation	19
2.3.1 Distributed Estimation	19
2.3.2 Sensing Capacity	20
III INFORMATION THEORY OVERVIEW	21
3.1 A Fundamental Problem In Information Theory	21
3.2 Extensions to the Fundamental Problem In Information Theory	25
IV WSN DESIGN TOOLS	27
4.1 Top-Level Design Metric	27
4.1.1 WSN Extension to Fundamental Problem - Perspective 2	28
4.1.2 WSN Extension to Fundamental Problem - Perspective 3	35
4.1.3 Positive Coding Theorems	37

4.1.4	Summary of Top-Level Design Metrics	38
4.2	WSN Sensing Efficiency Measure	39
4.2.1	Efficiency Measure Defined	39
4.2.2	Efficiency Analysis	41
4.2.3	Summary of Efficiency Analysis	44
4.3	New Sensor Design	45
V	SUMMARY	48
APPENDIX A	— TECHNICAL CONDITIONS FOR THE WSN MODEL	49
APPENDIX B	— PROOFS	52
APPENDIX C	— CALCULATIONS FOR THE EXAMPLE WSN	55
REFERENCES	59

LIST OF FIGURES

1	High-level functions in a WSN	2
2	WSN problem setup	3
3	Similarities between functionality of communication systems and functionality of WSNs.	4
4	Seven elements of the WSN model.	6
5	Techniques for creating a reproduction of a random variable.	13
6	Broadcast channel	15
7	Multiple access channel	16
8	Remote Source Coding	17
9	A fundamental problem in information theory	21
10	Perspective 1 on the fundamental problem in information theory . . .	23
11	Perspective 2 on the fundamental problem in information theory . . .	23
12	Perspective 3 on the fundamental problem in information theory . . .	24
13	Multi-terminal extension to the fundamental problem in information theory	26
14	WSN extension to the fundamental problem in information theory . .	28
15	Perspective 2 of the WSN extension to the fundamental problem in information theory	29
16	Adapting perspective 2 of the WSN extension to the fundamental problem in information theory	29
17	Information theoretic bound on the minimum number of sensors in the WSN for three values of δ and $\alpha = 1/100$	33
18	Performance comparison between the information theoretic bound developed in this work and the coverage approach implied by prior WSN research for $\delta = 1/2$ and $\alpha = 1/100$. Simulations used 10,000 time intervals and 100 random WSNs for each value of M	35
19	Perspective 3 of the WSN extension to the fundamental problem in information theory	35
20	Adapting perspective 3 of the WSN extension to the fundamental problem in information theory	36

21	Distortion efficiency of a WSN for $\delta = 1/2$ and two values of α	42
22	Distortion efficiency of a WSN for four values of δ and $\alpha = 1/100$. . .	43
23	Distortion efficiency of a WSN for $\delta = 1/2$, $\alpha = 1/100$, and four values of ϵ	45
24	Illustration of Traditional Source Coding	46

SUMMARY

We use tools and techniques from information theory to improve the design of Wireless Sensor Networks (WSNs). We do this by first developing a model for WSNs that is analogous to models of communication systems in information theory. In particular, we define the notion of WSN Coding, which is analogous to source coding from information theory, and the Collection Channel, which is analogous to a transport channel in information theory. We then use source coding theorems from information theory to develop three results that aid in WSN design. First, we propose a new top-level design metric for WSNs. Second, we develop an efficiency measure for the sensing process in a WSN. Finally, we use techniques from source coding schemes to suggest new designs for WSNs and the sensors they contain. We strive for tools that apply under the most general conditions possible so that designers can use them in any WSN. However, we also apply each tool to a specific example WSN illustrate the tool's value.

CHAPTER I

INTRODUCTION

This chapter provides background on Wireless Sensor Networks (WSNs) and our approach for improving their design. It also includes a description of the WSN problem we consider as well as our motivation for applying information-theoretic tools to the problem. Finally, we describe the WSN model used throughout the rest of the thesis and show how that model connects to elements traditionally used in information theory.

1.1 Background

WSNs consist of power-constrained sensors that are remotely deployed to measure certain features of a phenomena of interest. These sensors then transmit information back to a centralized location, called the sink, that will form an estimate of a desired quantity. For example, a WSN may have sensors deployed in a remote region of a forrest to collect information that will allow a sink to estimate if there is a fire. As with any technology, as WSN hardware and software mature, designers and technology developers will shift their focus from functionality to performance optimization. However, the nature of the WSN scenario makes this shift in focus critical for WSNs. In particular, optimizing WSN performance to maximize a WSN's lifetime is of utmost importance due to the fact that sensors are deployed to remote locations and therefore might not be replaced as their batteries run out. Because the sensors are power constrained, ensuring they are as efficient as possible and that the WSN uses the sensors as efficiently as possible is key to maximizing the WSN's lifetime. In this thesis, we develop three tools that WSN designers and technology developers can use as they optimize a WSN's performance to maximize its lifetime. First, we propose a

new top-level design metric for WSNs. Second, we develop an efficiency measure for the sensing process in a WSN. Finally, we use techniques from source coding schemes to suggest new designs for WSNs and the sensors they contain.

Ultimately, WSN optimization turns into a minimization problem with constraints coming from several parameters. From a cost perspective, we would like to minimize the operation and deployment costs of the WSN. From an energy-efficiency and network-lifetime perspective, we would like to minimize the number of sensors that are active at any one time. However, we must deploy enough sensors to successfully estimate the desired quantity, provide sufficient connectivity, and provide sufficient transport capacity, all with specified reliability.

We focus only on the sensing functionality of a WSN. Therefore, we will not consider other important WSN aspects such as sensors transmitting their measurements to the sink and the function the sink uses to make its estimate. We do this for two reasons. The first is that, as shown in Figure 1, the WSN can be viewed as a cascade of sensing, transmission, and estimation functions. Each of these elements contributes to the overall WSN performance, and the overall WSN performance will be limited by the weakest link in this chain. We would like to understand the limitations caused by the sensing functionality, so we assume optimal transmission and estimation functionality. The second reason to focus on the sensing functionality is that, as described in Section 2, transmission and estimation functionality have either been well studied in the information theory and WSN literature or closely represent problems in the information theory literature.

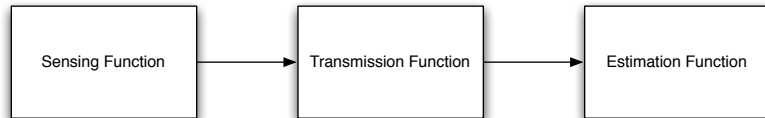


Figure 1: High-level functions in a WSN

We consider a WSN consisting of multiple sensors that are deployed in order to reproduce some feature(s) of a phenomenon at a sink node. The error in the sink node's estimate, called distortion, is calculated using fidelity criterion and must fall below a threshold specified by the network designer. For the purposes of this paper, the WSN's reproduction can be considered as a detection function or estimation function, and we refer to both cases as estimation. We use the term ground truth to refer to the features of the phenomenon that must be reproduced. This problem setup is shown in Figure 2.

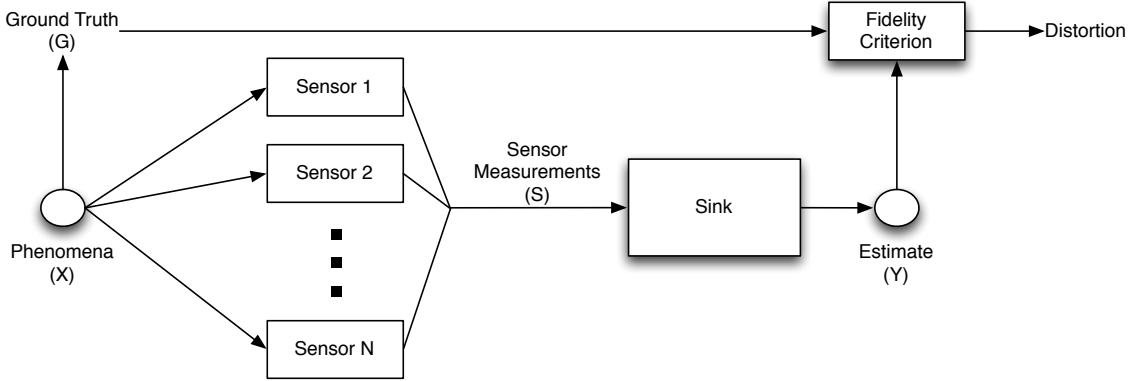
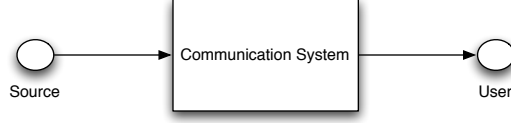


Figure 2: WSN problem setup

Our approach to developing these WSN optimization tools is motivated by the fact that, at a high level, a point-to-point communication system and a WSN have the same purpose: reproduce a random process at a remote location with some specified accuracy. As shown in Figure 3(a), in a communication system the random process is an arbitrary source that the communication system reproduces at the destination. As shown in Figure 3(b), in a WSN the random process is some feature(s) of a phenomenon that are sensed by the WSN and reproduced at the sink. Note that this purpose is much more specific than for traditional communication networks which are generally deployed for generic data transfer.



(a) Abstract model of a communication system's functionality.



(b) Abstract model of a WSN's functionality.

Figure 3: Similarities between functionality of communication systems and functionality of WSNs.

Information theory provides powerful tools for point-to-point communication system design (e.g. source codes and channel codes) and performance prediction (e.g. rate distortion bound, channel capacity) [18]. Therefore, from the perspective of this high-level problem description, information-theoretic tools provide an attractive method that can potentially be used for the design of and insight into WSNs.

Unfortunately the WSN problem is not identical to communication system problems where information-theoretic tools are traditionally employed. As a result, it is not immediately obvious how to apply information theoretic tools to the WSN problem.

For example, it might seem as if a WSN's sensing functionality requires a much more complicated description than a point-to-point communication system since a WSN potentially has many sensors taking noisy and possibly correlated measurements of the phenomenon. Considering each sensor's measurement individually would be problematic from an information-theoretic perspective because information-theoretic results are only available for specialized cases when there are multiple nodes [7].

Another example, many of the design parameters used to optimize communication system performance are fixed in a WSN once a certain sensor technology or

sensing application has been chosen. For instance, in source or channel coding for a communication system, designers optimize performance by controlling the mapping of input signal to the encoded signal that is output from the communication system. However, WSN designers cannot control the analogous parameters for a WSN's sensing functionality since picking a certain type of sensor will most certainly fix the mapping between features of an event and the sensors' measurements.

As a final example, the manner in which the sensors are placed through a geographic region is often fixed by the particular sensing application - the large number of sensors or remote geographic region makes random node placement the only feasible option. Again, in a traditional application such as quantization, defining quantization cells (which is analogous to choosing sensor locations) is a critical design parameter, and unfortunately we cannot usually optimize it in a WSN.

In spite of these challenges, this thesis focuses on finding a way to describe a WSN's functionality in the same way as a point-to-point communication system and then applying tools from point-to-point communication systems.

1.2 WSN Model

In this section, we describe our model of the WSN. Our development follows the approach from [22, 23]. Although [22, 23] propose ten elements in the WSN model, we only consider the seven elements shown in Figure 4: the *event area*, the *event model*, *ground truth*, the *placement model*, the *sensor measurements*, the *fidelity criterion*, and the *sink estimate*. Our goal is to use the most general models possible so that our results hold under the most general conditions. Therefore, we thoroughly define each element to illustrate the general nature of the model.

We model random processes as directly given random processes. In other words, random process $X = (x_0, x_1, \dots)$ will have induced probability space $(\Omega_X^\mathcal{T}, \mathcal{B}(\Omega_X)^\mathcal{T}, p_X)$ where \mathcal{T} is the index set, $x_i \in \Omega_X$, and $\mathcal{B}(\Omega_X)$ is the Borel σ -algebra of a set Ω_X .

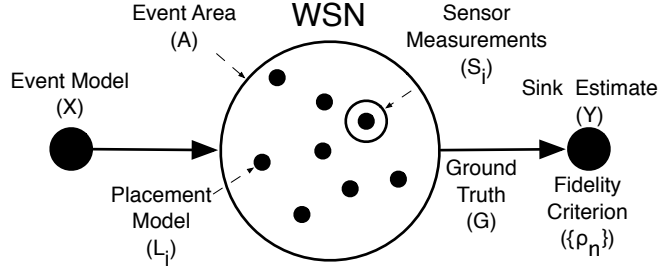


Figure 4: Seven elements of the WSN model.

When the meaning is clear from the context, we will often omit the superscript \mathcal{T} when referring to the random process to simplify notation.

Define the *event area*, A as the 2- or 3-dimensional region of space where the event must be sensed. Mathematically we write the event area as $A \subset \mathbb{R}^2$ or \mathbb{R}^3 . A might correspond to the area over which an event must be detected or the area over which a field must be estimated.

Define the *event model*, X , as a random process describing the event that the sensors will measure. Mathematically we write the event model as a collection random variables $\{X_t(\omega_E) : t \in \mathcal{T}\}$ defined on a common probability space $(\Omega_E, \mathcal{B}(\Omega_E), P_E)$ for index set \mathcal{T} . Typically, event models are broken down into two classes: point source events that occur at one point in a geographic area and field events that occur over all points in a subset of a geographic area. For a point source event, $x_t \in \Omega_{PS}$ where $\Omega_{PS} = (B, X_E, Y_E, Z_E) = (B, L_E)$ is a product space with B describing some characteristic of the event and L_E describing the event's position in the geographic area being sensed. For a field event, $x_t \in \Omega_{FE}$ where $\Omega_{FE} = \prod_{L_E} B_{L_E}$ is a product space with B_{L_E} describing some characteristic of the event at each point in the geographic area being sensed. Note that, if $B_{L_E} \neq 0$ for only a single point, the field event model is equivalent to the model of a point source event. Since both models are equivalent, we can maintain generality and also simplify notation by writing $x_t \in \Omega_X$ for all events.

Define *ground truth*, G , as the random process describing the quantity the sink needs to reconstruct. Ground truth is derived from the random process X by a mapping, \bar{f}_G where $\bar{f}_G : \Omega_X^T \rightarrow \Omega_G^T$. As examples of \bar{f}_G , consider two WSNs: a WSN where the sink needs to determine whether or not an event occurred, and a WSN where the sink needs to reconstruct a specific characteristic of a field event at every point in a region. In the first case, we let $g_t = 1_F(x_t)$ where 1_F is the indicator function and F is some subset of the geographic area being sensed. In the second case, we simply have $g_t = x_t$.

Even though the event model and ground truth seem like redundant elements, they are both necessary. The distinguishing factor between the event model and ground truth is that the event model describes the phenomenon the sensors measure while ground truth describes the quantity the sink must estimate. Sometimes these quantities are the same (e.g., estimating the value of a temperature field at each point in a given area using sensors that measure temperature at their location). However there are cases when the two quantities are different (e.g., estimating the average temperature over a given area using sensors that measure temperature at their location).

We place an important restriction on the mapping that generates G . We require that the G induced by \bar{f}_G be stationary or asymptotically mean stationary (AMS) depending on the fidelity criterion used for the estimate at the sink. For practical purposes, this requirement also means that X must be stationary or AMS as well. We require this so that source coding theorems will hold for G since those theorems will be key tool in the results that follow. As described in Appendix A, any realistic \bar{f}_G satisfies our requirement.

Define the *placement model* as the mechanism for placing the sensors throughout A . Mathematically we write sensor S_i 's location as the random variable $L_i : \Omega_L \rightarrow A$.

Define the *sensor measurements*, S_i , as the random processes describing the actual

sensor readings from measuring characteristics of the event. We assume that each sensor's measurement is derived from the event of interest X by a mapping, \bar{f}_S where $\bar{f}_S : \Omega_X^{\mathcal{T}} \rightarrow \Omega_{S_i}^{\mathcal{T}}$. For convenience, we let both S and $S^{(M)}$ denote the random process describing the all of the sensors' measurements (i.e., a time sample of S is a vector continuing the measurements of each sensor at that time). Just as we restricted \bar{f}_G , we restrict \bar{f}_S to those functions that produce a stationary or AMS S when X is stationary or AMS, respectively. As described in Appendix A, any realistic \bar{f}_S satisfies our requirement.

Given an event model X and a specified number of sensors, ground truth and the sensors' measurements will have an induced information rate, $\bar{I}(G; S^{(M)}) = I_{WSN}$. Therefore we may think of the WSN as being able to collect a certain amount of information about the event. This suggests the notion of a WSN's *collection channel* which is will be defined later. In addition, we can view the WSN as mapping a sample of the ground truth random process at each time instant to a set of sensor measurements. This suggests the notion of the WSN performing a type of coding of the ground truth random process. We term this *WSN coding* and will formally define it later.

Define the *sink estimate*, Y , as the random process describing the sink's estimate of ground truth. We assume that the sink's estimate is derived from the sensors' measurements by a mapping, \bar{f}_Y , where $\bar{f}_Y : \Omega_S^{\mathcal{T}} \rightarrow \Omega_Y^{\mathcal{T}}$. As mentioned earlier, we are only concerned with the sensing functionality of the WSN. Therefore we do not consider important WSN aspects such as how the sensors transmit their measurements back to the sink. Instead we assume the best case scenario in which the sink receives errorless copies of the sensor measurements which are synchronized in time. Furthermore, we assume that \bar{f}_Y is the optimal estimate the sink can make with respect to the fidelity criterion described in the next paragraph. Note that, since the sink's input is completely specified by the sensors' measurements, the WSN forms a

Markov Chain $G_t \rightarrow S_t^{(N)} \rightarrow Y_t$. We will use this fact later in proving Theorem 1.

Define the *fidelity criterion*, $\{\rho_t\}$, as the measures of how accurate the sink has estimated ground truth. This is the traditional fidelity criterion from information theory which we denote $\{\rho_t(g^t, y^t)\}$. As with \bar{f}_G and \bar{f}_S , we restrict $\{\rho_t\}$ to be functions for which source coding theorems hold. For example, if G is AMS, we require ρ_t to be an additive function. And, if G is stationary, we require ρ_t to be a sub-additive function. Again this restriction is mild since most fidelity criterion of interest satisfy this requirement.

Given X , G , S , and $\{\rho_t\}$ we can determine the \bar{f}_Y that generates the smallest expected distortion, which we denote D_{WSN} .

We have done two things by defining the WSN model this way. First, we have ensured that source coding theorems hold for a WSN with the given G , S , and $\{\rho_t\}$. We will use this fact in Section 4.1 to show that it is not possible for the sink to estimate the event with the required accuracy unless the I_{WSN} exceeds a specified value. Second, we have ensured the WSN can be compared to traditional source codes from information theory using I_{WSN} and D_{WSN} . We will use this fact in Section 4.2 to calculate the efficiency of the sensing process in a WSN.

1.3 WSN Coding and the Collection Channel

Theorems from information theory are typically defined in terms of either channels or codes. To ensure consistency with information theory literature, we define a WSN's sensing functionality in both of these terms.

In information theory, traditionally a mapping from input random variable to output random variable is called a *channel* if the mapping is random and called a *code* if the mapping is deterministic. From this perspective, the mapping between ground truth and sensor measurements can be either a channel or a code depending on \bar{f}_S - although for most realistic WSN models this mapping is random and would

be considered a channel. However, information theoretic characterizations of a channel are useful because they describe how a signal should be modified before being transmitted by a communication system so that the signal is reliably received at the communication system’s remote location. These characterizations are potentially not as useful for describing the sensing functionality of a WSN because the event cannot be modified before it is measured by the sensors. On the other hand, information theoretic characterizations of a code are useful because they describe the minimum amount of information a code must provide to reliably reconstruct a signal. In the WSN scenario, this is similar to characterizing the sensor measurements needed to reliably reconstruct an event. From this perspective, a WSN’s sensing functionality more closely resembles a code. As noted in [18], a code is a special case of a channel. Therefore we compromise and use both terms for a WSN. We use the more general term *collection channel* when the WSN has a fixed number of sensors and we are considering general classes of events. This will allow us to ask, “Given a WSN with a specified number of sensors, what types of events can be estimated?” We will use the term *WSN Coding* when the number of sensors is a design parameter and we are considering a specific event or class of events. This will allow us to ask, “Given an event, what type of WSN is needed?” As above, our development of these terms follows the approach from [22, 23].

1.3.1 Collection Channel

Formally, a channel is simply defined as a family of regular conditional probability measures on a measurable space. We adopt this definition while taking into account the parameters from our WSN model. Clearly, the mapping of event features to sensor measurements depends on M , the number of sensors in the WSN, how the sensors are placed throughout the event area, and the event area itself. Using these variables, we can formally define a collection channel.

Definition: Following the notation in [18], the *collection channel*, $[\Omega_X, \nu, \Omega_S]$, with input alphabet Ω_X and output alphabet Ω_S is a family of regular conditional probability measures $\{\nu_{x|L,A} : x \in \Omega_X^T\}$ on the measurable space $(\Omega_S^T, \mathcal{B}(\Omega_S^T))$.

Two primary concerns when considering different types of channels are whether they are stationary and whether they are ergodic, since these conditions determine whether or not theorems from information theory hold. As shown in Appendix A, most realistic L and \bar{f}_S result in a collection channel that is primitive. It is well known that primitive channels are stationary (AMS, ergodic) with respect to any stationary (AMS, ergodic) source X [31]. Therefore, theorems from information theory will hold for most realistic collection channels.

1.3.2 WSN Coding

Formally, given a source random process, a code is simply defined as a family of mappings from the source process alphabet to a code alphabet. We adopt this definition while taking into account the parameters from our WSN model. As mentioned earlier, we can consider the sensors' measurements as a mapping from the ground truth random process to the random process describing the sensors' outputs. If we let \mathcal{C}_{WSN} denote the WSN code we can write this as $\mathcal{C}_{WSN} : \Omega_G \rightarrow \Omega_{S(M)}$. The key difference between a WSN code and a traditional code from information theory is that a WSN code employs a probabilistic mapping from input process to output process while traditional information theory codes use a deterministic mapping.

Definition: Given a WSN with M sensors, and specified A , L , and \bar{f}_S , a *WSN code* is a mapping from a ground truth random process to M sensor measurements. Mathematically the WSN code is written $\mathcal{C}_{WSN} : \Omega_G \rightarrow \Omega_{S(M)}$.

Using this definition, we could use concepts from source coding theory to define the WSN code's coding rate in terms of the number of bits per source symbol. Clearly this coding rate would be directly related to the number of sensors in the WSN. Because

of this relationship, we will instead define the rate of the WSN code in terms of the number of sensor measurements per source symbol. In this case, the code's rate is simply the number of sensors, M . As in traditional information theory we would like to examine the tradeoffs between code rate and achievable distortion for the WSN code.

1.4 Positive Coding Theorems vs. Negative Coding Theorems

As a final connection to information theory, we note that channels and codes usually have a positive coding theorem as well as a negative coding theorem. Positive coding theorems provide conditions under which a specified channel or coding performance can be achieved. For example, a positive coding theorem might state the minimum bit rate at which the required distortion can be achieved. On the other hand, negative coding theorems give conditions under which the channel or coding performance cannot be achieved. For example, a negative coding theorem might state that the required distortion cannot be achieved under a certain bit rate.

From a WSN perspective, we would like to prove both a positive and negative theorem for the collection channel as well as a positive and negative theorem for WSN coding. Unfortunately, as will be seen later, the positive theorems can only be provided when considering very small classes of WSNs. This violates our goal of developing tools for general classes of WSNs, and as a result, positive theorems not useful for WSN design. Therefore, this thesis will primarily focus on proving negative coding theorems.

CHAPTER II

RELATED WORK

As stated above, a WSN can be viewed as a mechanism to represent one random process (ground truth) with another random process (the estimate at the sink) that is a sufficiently accurate estimate of the original random process. This abstract description closely resembles many other problems in engineering. In fact, a communication system will by itself encounter many of these problems including sampling, quantization, source coding, and channel coding. As shown in Figure 5, the problem of creating a reproduction of a random process can be broken down into two main categories: asymptotic coding (also called source coding with a fidelity criterion) and one-shot coding (of which quantization and sampling are subsets) [20]. The primary difference between the two approaches is that asymptotic coding focuses on coding of typical sequences in a random process while one-shot coding focuses on coding finite blocks of samples from a random process [20].

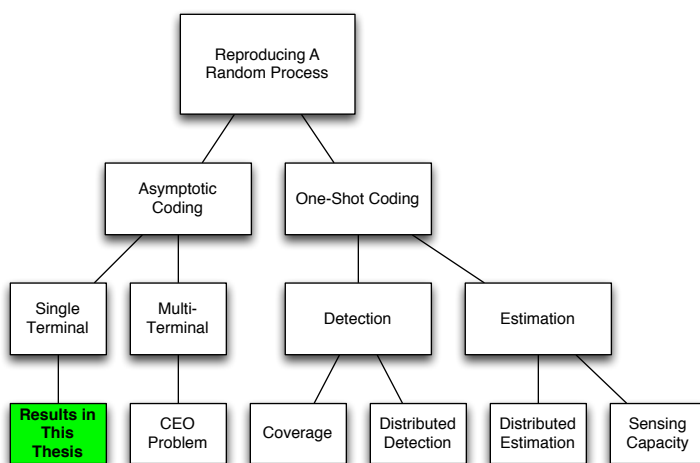


Figure 5: Techniques for creating a reproduction of a random variable.

Within asymptotic coding, there are two major classes of problems: single-terminal coding in which one node creates the representation of the random process, and multi-terminal coding in which multiple nodes create representations of the random process. Within one-shot coding, two major classes of problems are particularly relevant to WSNs: detection and estimation. As noted in [37], even though detection and estimation are traditionally considered separately, the estimation problem is actually a version of the detection problem. The difference is that the two problems have a different emphasis; estimation usually considers a large reproduction alphabet and accuracy specified by mean-square error (MSE) while detection usually considers small reproduction alphabets and accuracy specified by probability of error.

Prior research has tailored existing results from these problem classes to consider characteristics specific to WSNs. However, the primary focus has been on one-shot coding. From a practical perspective, this means that most efforts have implicitly assumed that WSNs report a limited number of measurements to the sink regardless of how long the WSN is deployed. In fact, most prior work implies that only one measurement is reported to the sink. Just as in traditional communication systems, this assumption can be overly restrictive. Multi-terminal asymptotic coding results have also been applied to WSNs but typically for the problem of transmitting sensors' measurements to the sink.

Below we discuss five representative problems from the framework shown in Figure 5. Each of these problems has been specifically tailored to consider WSN characteristics. These problems are the CEO problem, coverage, distributed detection, distributed estimation, and sensing capacity. For each problem we describe the basic model used and the questions typically considered. In the discussion, we occasionally deviate from the notation in the original work for easier comparison to the model proposed in Section 1.2.

2.1 Multi-Terminal Coding

The WSN problem is very similar to problems from multi-terminal information theory. One such problem is the broadcast channel shown in Figure 6. In the source-coding perspective of this problem (vice the channel-coding perspective), a single source, X , must be encoded and transmitted to M number of destinations. Information theory has been used to calculate how much information must be transmitted between the source node and destination nodes in order for the destination nodes to accurately reproduce the source signal [13]. At first glance, this is similar to the problem of determining how much information the WSN must collect about the event of interest. However, as seen in Figure 6, the broadcast channel assumes that the source signal can be encoded before it is transmitted to the destinations. In the WSN problem, the source signal corresponds to the random process describing ground truth, and the destination nodes correspond to the sensors. Due to the nature of the problem, ground truth cannot be encoded before it is measured at each sensor. Therefore the broadcast channel results do not apply to our scenario where we focus on the WSN's sensing functionality. The channel-capacity perspective of this problem examines the maximum amount of information that can be transmitted to the destinations, and this is not applicable to the sensing functionality of the WSN.

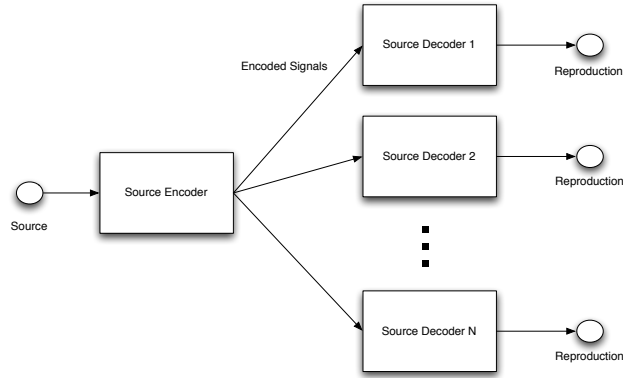


Figure 6: Broadcast channel

Another multi-terminal information-theory problem is the multiple access channel shown in Figure 7. In the source-coding perspective of this problem, M correlated sources, S_i , are independently encoded at multiple encoders and decoded together at a single decoder. Slepian-Wolf coding has been used to determine how much information must be transmitted between the source nodes and the destination in order for the destination to accurately reproduce all of the source signals [13]. At first glance, this is similar to the WSN problem of transmitting sensor measurements to the sink node. However, as seen in Figure 7, the multiple access problem assumes that the signal to be reproduced at the destination node is exactly the signal contained at each encoder. In the WSN problem, the encoders correspond to the sensors, and the reproduction corresponds to the sink. As described in the problem setup, we do not want to reproduce the sensor measurements. We want to reproduce ground truth, and it is not always clear that having all the sensor measurements will allow us to produce an accurate reproduction of ground truth. Therefore, the multiple access channel results do not apply to our scenario where we focus on the WSN's sensing functionality. It should be noted that the multiple access results can perhaps be applied to the transmission function in the WSN since the transmission functionality is focused on reproducing the sensors' measurements at the sink.

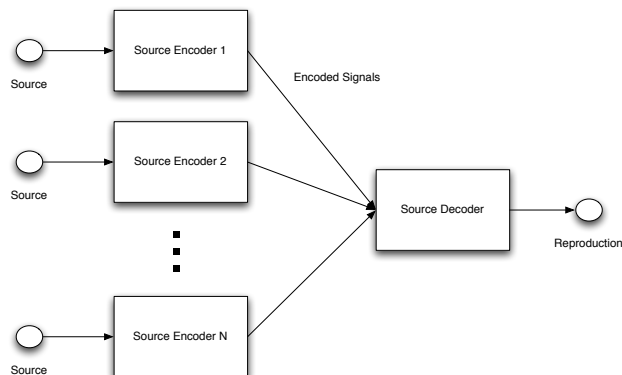


Figure 7: Multiple access channel

If we consider a multiple access channel with only one source encoder that is used

to encode a noisy version of a source signal, we get the remote source coding problem shown in Figure 8 [6]. Remote source coding tells us the distortion and bit-rate performance limits that the system can achieve when trying to reproduce the original source signal. At first glance, this problem strongly resembles the WSN problem since the WSN’s sensors take noisy measurements of an event. However, since we consider the WSN’s sensing functionality, the error is not a separate component. Therefore, these results do not apply to our scenario.

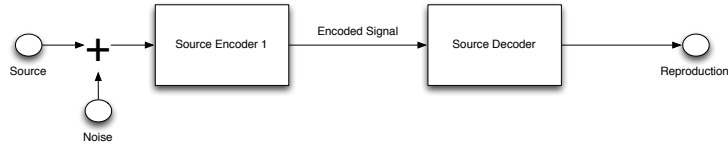


Figure 8: Remote Source Coding

Finally, The CEO problem is a multi-terminal information-theory problem that has been well studied in information theory literature [5]. In the CEO problem, a phenomenon X is observed by M agents (or sensors in our case). These agents report their noisy measurements, S_i , to the CEO who forms an estimate, Y , of the phenomenon. The objective is to minimize both the total transmission rate between the agents and the CEO, and the distortion between the CEO’s estimate and the phenomenon. The block diagram for the CEO problem is similar to our WSN setup in Figure 2. In fact, the model used for the sensors in the CEO Problem can be considered a special case of the model we consider. This area of research typically examines the tradeoff between the transmission rate and distortion as the number of sensors goes to infinity. As mentioned above, this area of research typically does not consider sensor placement or the shape of the sensors’ sensing regions.

In [16], the model used in the CEO problem has been generalized and used to derive results similar to our results in Section 4.1. The authors consider a specific WSN where each sensor’s measurement is an additive noise process $S_i = W_i + \alpha_i X_i$

and the phenomenon X_i and additive noise W_i are modeled by IID circularly complex Gaussian random processes. The authors then use source and channel coding theorems from information theory to determine bounds on the sink's distortion as a function of the number of sensors M . In Section 4.1 we seek similar results but for more general classes of sensor networks and phenomenon.

2.2 Detection

2.2.1 WSN Coverage

The WSN coverage problem has been well studied in networking literature [1, 4, 10, 19]. In its most basic form [19], the WSN coverage problem considers a WSN consisting of M sensors, $\{S_i : i = 0, \dots, M - 1\}$ deployed over a region of space. Each sensor has a sensing area, and it is assumed that any event occurring inside of a sensor's sensing area will be detected. A region is said to be covered if certain, or perhaps all, subsets of the region are included in at least one sensor's sensing area. This area of research examines how many sensors are needed to cover a given area with a specified probability. The basic results have been extended for new definitions of coverage including exposure [30], K-coverage [21], probabilistic coverage [2, 3], and information coverage [39].

2.2.2 Distributed Detection

The distributed detection problem is a statistical hypothesis testing problem that has been well studied in signal processing literature [37]. In the basic distributed detection problem, a random variable X can satisfy one of M hypotheses. As in the CEO problem, M sensors take a noisy measurement of X denoted $S_0(X), \dots, S_{M-1}(X)$. Each sensor then uses a local decision rule on its measurement to select a finite-valued message, S'_i , which is sent to the fusion center (or in our case, the sink). The fusion center then forms an estimate, $Y = g(S'_0, \dots, S'_{M-1})$, of the original hypothesis using a global decision rule. Performance of the fusion center is usually determined by

the probability of error of the global estimate. This area of research examines the relationships of various performance metrics for differing local and global decision rules. The basic results have been tailored to features specific to WSNs [12].

2.3 Estimation

2.3.1 Distributed Estimation

The distributed estimation problem is a quantization problem that has also been well studied in signal processing literature [27, 40]. In the basic distributed estimation problem, M sensors take a noisy measurement, $S_0(X), \dots, S_{M-1}(X)$, of the random variable X . Each sensor then quantizes its measurement into a finite set of values. The sensors send the quantized values, S'_0, \dots, S'_{M-1} , to a fusion center that forms an estimate, $Y = g(S'_0, \dots, S'_{M-1})$, of the random variable. The performance of the fusion center is typically determined in terms of the MSE between the fusion center's estimate and the actual value. This area of research typically examines how to design the quantizers and the estimation function at the fusion center. The basic results have also been tailored to features specific to WSNs [28]. In fact, [28, 29] imply that there exists a sensor minimum similar to the minimum we derive in Section 4.1.

We note that our model in Section 1.2 can be tailored to represent many distributed estimation problems. For example, in most WSN research, sensors are modeled as being able to measure an event only if the event takes place within a circular area centered around the sensor. If we assume this circular area has radius much smaller than the spatial bandwidth of the event, our WSN model represents the basic distributed estimation problem [27]. If then we assume that L_i can be chosen so that the sensors' measurements form a Riesz basis, the sensors' measurements can be seen as projections on to a random space, and our WSN model could also characterize a sampling problem [38]. If we further restrict the placement model to be deterministic and uniform, our WSN model constitutes a traditional Nyquist sampling problem

[38]. If instead we restrict the placement model to be deterministic and optimized to cover the area, our WSN model represents a vector quantization problem [17].

2.3.2 Sensing Capacity

The sensing capacity problem has been studied in information theory literature [33, 34]. In the basic sensing capacity problem, a phenomenon is to be estimated at L points and described by an L -dimensional random vector, (X_0, \dots, X_{L-1}) . The sensor network consists of M sensors that can each observe a subset of the L points so that each sensor's measurement is a combination of measurements from different points. The subset of points that a sensor can observe is determined by the geography of the WSN. The sensors' noisy measurements, $S_0(X), \dots, S_{M-1}(X)$ are then relayed to the sink so that it can form an estimate, $Y = g(S_0, \dots, S_{M-1})$, of the original L -vector describing the phenomenon. The notion of sensing capacity is defined as the ratio L/M of points to sensors that a given sensor network can observe. This area of research examines how sensing capacity changes with different WSNs. Note that this definition of sensing capacity is somewhat different than our notion of collection capacity since sensing capacity is described in terms of points per sensor while collection capacity, like transmission capacity, is described in terms of bits. Furthermore, the sensing capacity problem is different than the WSN problem we consider. Therefore, sensing capacity results are not applicable to our problem.

CHAPTER III

INFORMATION THEORY OVERVIEW

Since information theory provides the key tools for our results in the next chapter, we describe a fundamental problem in information theory that will be used extensively in our results. Our description is given from perspectives that will allow us to apply the results to WSNs.

3.1 A Fundamental Problem In Information Theory

Consider the setup shown in Figure 9. Since variations of this figure will be used extensively in this chapter as well as the next, we pause to explain the coloring scheme. Elements in blue will be fixed for a given problem. Elements in purple will be the answer to the question posed by a particular problem. Elements in red will be variable for a given problem in the sense that each specific value for these elements will correspond to a value for the problem's answer.

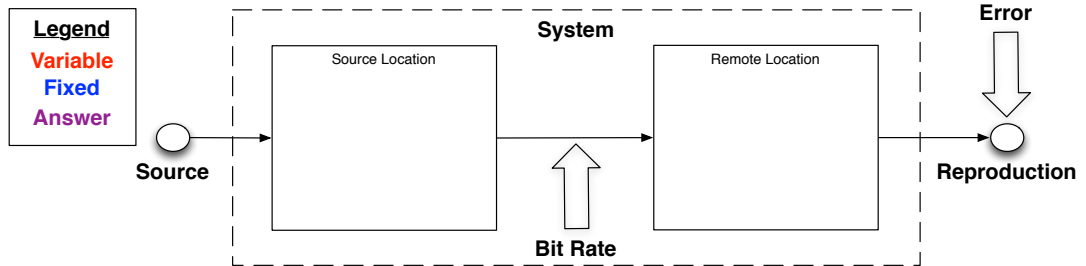


Figure 9: A fundamental problem in information theory

We now describe the different elements of the problem. There is a source random process that must be reproduced at a remote location. A system will be used to create this reproduction. The system has an associated bit rate that describes the maximum amount of information that can be transmitted between the source location and the

remote location. Finally, there is a maximum long-term average error allowed in the reproduction. A fundamental problem in information theory is to determine the relationships between these four elements. We will describe three different perspectives on this problem that each describe a different relationship.

Although information theory techniques strive to determine relationships between these elements under the most general conditions possible, there are two key restrictions. First, the results only apply in certain problem formulations. In particular, we must consider an infinite-length sample sequence of the random process and we must also consider the long-term average error. This restriction effectively means that we are in the asymptotic coding regime from Figure 5. Second, certain technical conditions must hold. In particular, we require that the source signal and the signal transmitted to the remote location both be a stationary or asymptotically mean stationary (AMS) random process, and, if the source signal is stationary (AMS), the fidelity criterion must be a sub-additive (additive) function. We make this restriction so that information-theoretic quantities such as mutual information exist for the infinite-length sequences we consider.

Our first perspective on the fundamental problem, which is shown in Figure 10, is perhaps the most natural for a system designer. In Perspective 1, we are given a source random process, and we consider ALL systems with bit rate less than a specified value. It is important to note that system designers can be as creative as possible here, and they can design any system as long as the bit-rate restriction holds. Source coding theorems from information theory then describe the best possible error performance that ANY of these systems can achieve. This is a remarkable result considering how few restrictions we have placed on the systems under consideration. The performance bound from these source coding theorems then provides system designers with a performance goal for any system that they design as well as a measure of “how good” their system is.

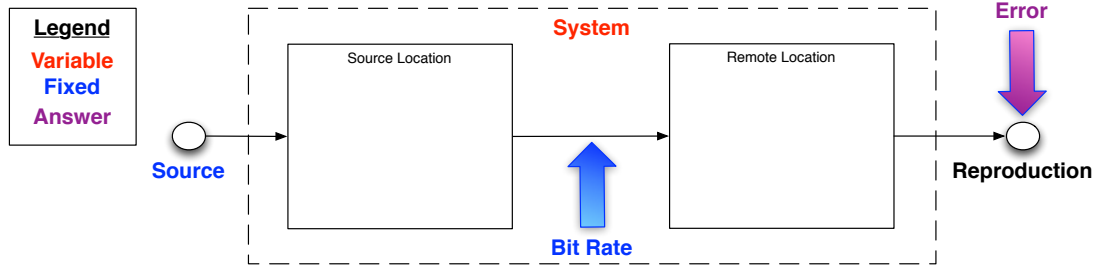


Figure 10: Perspective 1 on the fundamental problem in information theory

Even though the Perspective 1 on the fundamental problem is the most natural for a system designer, most results from information theory are focused on a second perspective shown in Figure 11. In Perspective 2, we are given a source random process, and we consider ALL systems with an error less than a specified value. As before, system designers can be as creative as possible as long as the error restriction holds. Source coding theorems from information theory then describe the smallest possible bit rate that ANY of these systems can achieve. Again, this is a remarkable result considering the general class of systems it applies to, and the performance bound can be used as a design goal for individual systems.

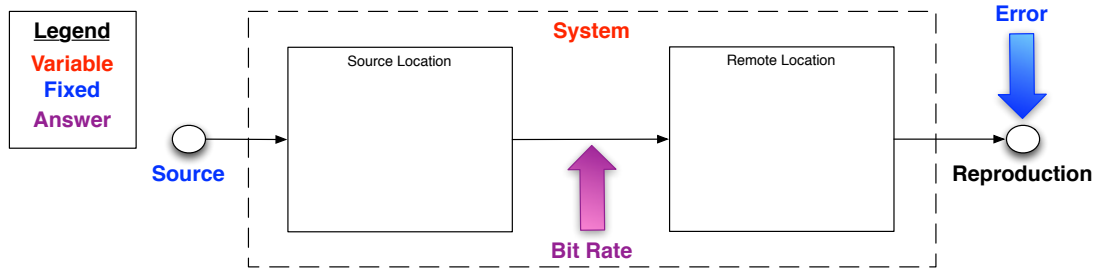


Figure 11: Perspective 2 on the fundamental problem in information theory

Perspectives 1 and 2 on the fundamental problem are closely related; in fact, they are the inverse of each other. Traditionally, Perspective 1 is referred to as a distortion-rate problem, and the performance bounds are given by the distortion-rate function, $D(R)$. The distortion-rate function describes, for a given bit rate R , the minimum possible distortion the system can have. Perspective 2 is referred to as a

rate-distortion problem, and the performance bounds are given by the rate-distortion function, $R(D)$. The rate-distortion function describes, for a given distortion D , the minimum possible bit rate in the system. The rate-distortion function can be formally written

$$R(D) = \limsup_{t \rightarrow \infty} \inf_{q_t \in Q_{t,D}} t^{-1} I(G^t; Y^t), \quad (1)$$

where q_t is the conditional probability of Y^t given G^t , $I(G^t; Y^t)$ is the mutual information between G^t and Y^t when their conditional distribution is given by q_t , and $Q_{t,D} = \{q_t : t^{-1} E[\rho_t(g^t, y^t)] \leq D\}$ is the set of all q_t with average expected distortion less than D .

Since most of the theory has developed around Perspective 2, we will use the rate-distortion function extensively to prove the WSN Coding results in Section 4.1. We will use both the distortion-rate function and the rate-distortion function in Section 4.2 to develop an efficiency measure.

Our final perspective on the fundamental problem is shown in Figure 12. In Perspective 3, we are given a system with an associated bit rate, and we consider a specified maximum error in the reproduction. Source coding theorems from information theory then describe the classes of all random processes the system can reproduce with the required error. We will use this perspective to prove the Collection Channel result in Section 4.1.

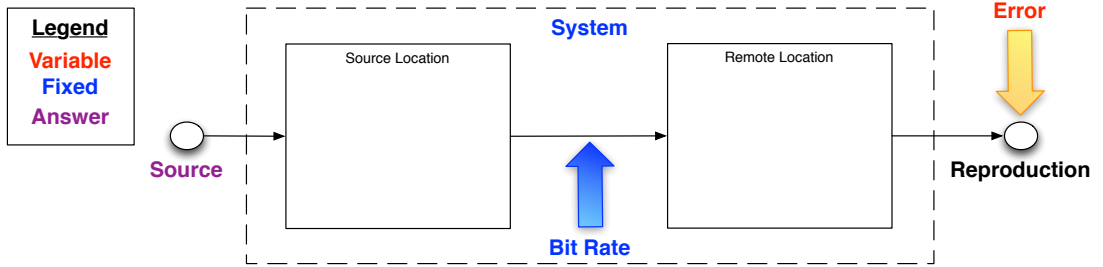


Figure 12: Perspective 3 on the fundamental problem in information theory

We now give a concrete example to illustrate these different perspectives. We

let the source random process be the flip of a coin. Therefore an example source sequence might be $X = H, T, T, H, T, H, H, H, \dots$. The system will encode the source sequence into a signal that is transmitted to the remote location. For example, the system might take three source symbols such and map them into a binary string that is transmitted to the remote location (e.g., $X = H, T, T$ is mapped into the binary string 0, 1). The remote location then estimates the original source sequence from the binary sequence. The error measurement judging the quality of the reproduction is calculated as follows: if the source is H and the reproduction is T or if the source is T and the reproduction is H , the error is equal to 1; otherwise the error is equal to 0.

Using this setup, Perspective 1 tells us that ANY system with bit rate smaller than R must have average error greater than $D(R)$, where $D(R)$ is the inverse of equation (2). Perspective 2 tells us that ANY system with average error less than D has bit rate greater than

$$R(D) = -\log\left(\frac{1}{2}\right) + D \log(D) + (1 - D) \log(1 - D). \quad (2)$$

Finally, Perspective 3 says that the system for encoding the original source process can transmit ANY random process whose rate-distortion function, $R(D)$, is greater than the system's associated bit rate, R , for any given distortion, D .

3.2 Extensions to the Fundamental Problem In Information Theory

Information theory is a very powerful tool because the theorems relating the four elements in the fundamental problem hold under very general conditions for the source, system, bit rate, and error measure. However, many times the fundamental problem is extended by imposing certain restrictions on the structure of the system under consideration in order to determine if the theorems from information theory still hold. We call imposing a specific structure on the system under consideration an “extension”

to the fundamental problem.

As a concrete example of an extension, consider Figure 13. Here we impose a structure on the system at the source location. In particular, we assume the source location receives the source signal at multiple encoders, and each encoder separately sends a coded version of the signal to the remote location. In fact, this is the multiple access channel problem described in Section 2.1. For this example, we adopt Perspective 2 so that, given the source random process, we would like to know the smallest bit rate for any system with the specified encoder structure at the source location and the specified error performance.

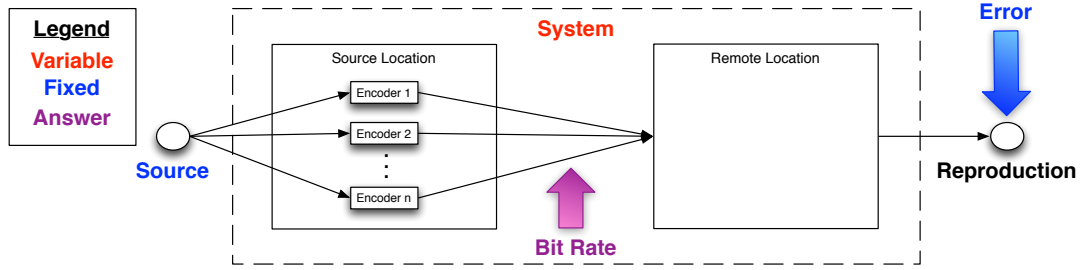


Figure 13: Multi-terminal extension to the fundamental problem in information theory

As described in the next chapter, the idea of extending the fundamental problem will be key to applying information-theoretic results to WSNs.

CHAPTER IV

WSN DESIGN TOOLS

In this chapter, we develop three results that will aid in WSN design. First, we propose a new top-level design metric for WSNs. Second, we develop an efficiency measure for the sensing process in a WSN. Finally, we use techniques from source coding schemes to suggest new designs for WSNs and the sensors they contain. We strive for tools that apply under the most general conditions possible so that designers can use them in any WSN. However, we also provide a specific example for each tool to illustrate its value.

4.1 Top-Level Design Metric

As WSN hardware and software mature, practical questions related to WSN operation and deployment will become more important - especially as they relate to the number of sensors the WSN needs to have. One of the most relevant questions is, “How many sensor nodes must successfully report their measurements for the WSN to operate as desired?” Looking at the same issue from a different perspective, we would also like to answer the question, “What kinds of phenomena can a previously deployed WSN accurately detect?” Answers to these questions impact many aspects of WSN design including architecture [11, 15, 36], deployment strategies [26, 41], protocol design [8, 14], and power management [9, 32]. Furthermore, the requirement to minimize a WSN’s total ownership cost means that the number of sensors in a WSN could provide a useful comparison when evaluating candidate sensor and vendor technologies for implementing the WSN.

To answer the above questions from the perspective of sensing functionality, we follow the approach in [22, 23] and extend the fundamental problem in information

theory as follows. First, we let the source random process be the ground truth to be estimated. Then, we restrict the source location to be a WSN that measures the features of the phenomena. Finally, we assume perfect transmission to and optimal reconstruction at the remote location, which represents the sink. This problem setup is shown in Figure 14. An important note regarding this extension is that we want to impose as little structure as possible on the WSN. That way, any results are as broadly applicable as possible.

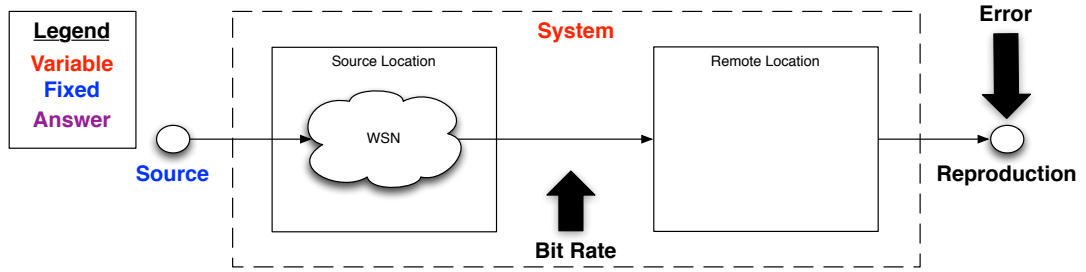


Figure 14: WSN extension to the fundamental problem in information theory

In the remainder of this section, we would like to find out if there are results from information theory that can be applied to this problem setup in order to design a more optimal WSN.

4.1.1 WSN Extension to Fundamental Problem - Perspective 2

In order to answer the above question regarding the minimum number of sensors, we take Perspective 2 on the WSN extension to the fundamental problem in information theory. This problem setup is shown in Figure 15.

As mentioned in Section 3.1, source coding theorems from Perspective 2 of the fundamental problem in information theory tell us that, to achieve a certain distortion D , the system must have bit rate greater than the rate-distortion function, $R(D)$.

As shown in Figure 16, even if we don't assume any structure on the WSN, we know that the bit rate of useful information being sent to the sink is limited by I_{WSN} , the information that the WSN collects about the source.

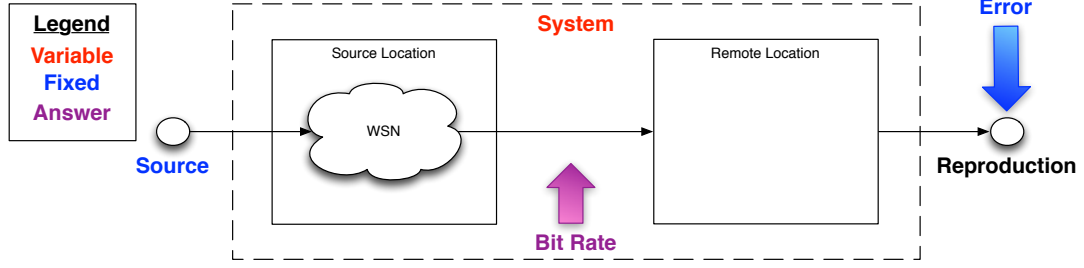


Figure 15: Perspective 2 of the WSN extension to the fundamental problem in information theory

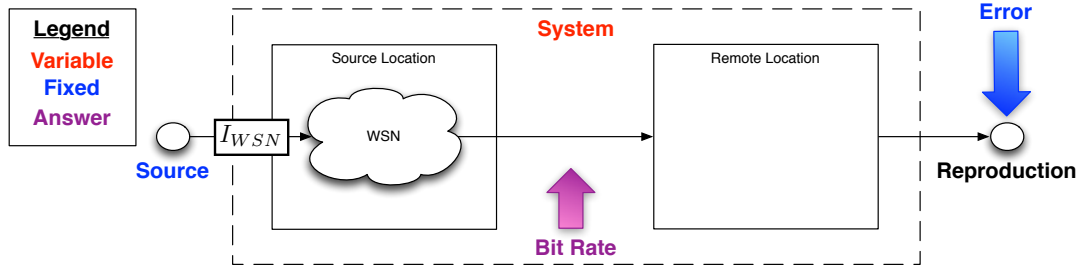


Figure 16: Adapting perspective 2 of the WSN extension to the fundamental problem in information theory

We can use this fact to make $R(D)$ a top-level design metric to determine the minimum number of sensors the WSN must contain. Below this minimum, the WSN will not be able to reproduce ground truth with the required accuracy. This result is formally stated in the following theorem and proved in Appendix B.

Theorem 1 (Negative WSN Coding Theorem) *Consider a WSN coding scheme $S : \Omega_X \rightarrow \Omega_{S(M)}$ with coding rate (in terms of number of sensor measurements per source symbol) M . If the code induces a mutual information rate between ground truth and sensor measurements less than the rate distortion function for ground truth and the sink's estimate, the sink's estimate will not have the required accuracy. In other words, $I_{WSN} = \bar{I}(G; S^{(M)}) \leq R(D)$ implies $\lim_{t \rightarrow \infty} t^{-1} E[\rho_t(g^t, y^t)] \geq D$.*

We can then use the Negative WSN Coding Theorem to prove the key result for WSN designers. This result states that there is a critical number of sensors in the WSN. If the WSN has fewer than this number of sensors, the WSN will not be able

to perform as required. This result is formally stated in the following lemma and proved in Appendix B.

Lemma 1 (Minimum Bound Lemma) *Let M_{min} be the largest M such that*

$$\bar{I}(G; S^{(M)}) \leq R(D).$$

For any $M < M_{min}$, the sink will not be able to estimate ground truth with the required average distortion. In other words, $M < M_{min}$ implies $\lim_{t \rightarrow \infty} t^{-1} E[\rho_t(g^t, y^t)] \geq D$.

4.1.1.1 WSN Coding Example

We use an example to illustrate the efficiencies that result from using the information theoretic measures from the above results as top-level design metrics. We assume the WSN's objective is to determine whether a specific event occurs in a given area. This means the WSN makes a binary decision over a number of time intervals, and the sink's resulting output will be a binary random process. Furthermore, we assume the WSN uses sensors that can perfectly detect the event if the event takes place in a given area around the sensor. Even though this model is simple, it is commonly used in WSN literature [25].

First we define the parameters for this example using the WSN model in Section 1.2.

Let the event model, X , be an independently and identically distributed (i.i.d.) point source so that the distribution of X_t equals X for all t . Furthermore, assume that the location of X is a random variable that is uniformly spread over the event area, A .

Since the WSN's function is to detect an event, we can describe *ground truth* as whether or not the amplitude of the event at time t , has crossed a threshold. In other

words, $\bar{f}_G : \Omega_X \rightarrow \{0, 1\}$ so that

$$G_t = \bar{f}_G(\omega_x) = \bar{f}_G(b, l_E) = \begin{cases} 1 & \text{if } b > C \\ 0 & \text{if } b \leq C \end{cases} \quad (3)$$

where $b \in B$ is the event's amplitude and $l_E \in A$ is the event's location.

This induces a distribution for ground truth with probabilities

$$P_{G_t}(G_t = 1) = P_X(b > C) = \delta$$

$$P_{G_t}(G_t = 0) = 1 - \delta.$$

Because $\{X_t\}$ is i.i.d. with respect to t , so is $\{G_t\}$. Therefore, we can describe ground truth as a binary i.i.d. random process G where $Pr(g_t = 1) = \delta$.

Let the sensing model, S_i , be as follows. If the event occurs within a circular region centered around a sensor and the event's amplitude is greater than a certain value, the sensor will perfectly detect that the event occurred. Let A_i denote the circular region around sensor. In other words, $\bar{f}_S : \Omega_X \rightarrow \{0, 1\}$ so that

$$S_{i,t} = \bar{f}_S(\omega_x) = \bar{f}_S(b, l_E) = \begin{cases} 1 & \text{if } b > C, \text{ and } l_E \in A_i \\ 0 & \text{otherwise} \end{cases}. \quad (4)$$

We assume the placement model L is such that the sensors are placed randomly, uniformly, and independently throughout the event area. Therefore $\{L_i\}$ are IID with respect to i . Furthermore, we assume that A_i is small relative to the event area, and let $\alpha = A_i/A$. With these assumptions we know that $Pr(l_E \in A_i) = \alpha$.

As shown in Appendix C, the WSN's minimum distortion, D_{WSN} , and mutual information, I_{WSN} , are written

$$D_{WSN} = \delta(1 - \alpha)^M \quad (5)$$

and

$$I_{WSN} = H_b(\delta) - (1 - \delta + D_{WSN}) H_b\left(\frac{1}{1 + \frac{D_{WSN}}{1 - \delta}}\right) \quad (6)$$

where M is the number of sensors in the WSN.

Let the *fidelity criterion* be the single-letter Hamming distance. Using the well-known result for an IID binary source with probabilities δ and $1 - \delta$ and fidelity criterion given by the single-letter Hamming distance [6], we can write the rate-distortion function for ground truth as

$$R(D) = H(G) - H_b(D) = H_b(\delta) - H_b(D) \quad (7)$$

where $H_b(q) = -q \log(q) - (1 - q) \log(1 - q)$ is the binary entropy function.

Lemma 1 states that the minimum number of sensors in the WSN is given by the smallest M such that $I(G; S^{(M)}) \geq R(D)$. Using the results from (6) and (7), this means that

$$H_b(\delta) - (1 - \delta + D_{WSN}) H_b\left(\frac{1}{1 + \frac{D_{WSN}}{1 - \delta}}\right) \geq H_b(\delta) - H_b(D). \quad (8)$$

In other words, our bound is given by the smallest M such that

$$H_b(D) \geq (1 - \delta + D_{WSN}) H_b\left(\frac{1}{1 + \frac{D_{WSN}}{1 - \delta}}\right). \quad (9)$$

Since we have parameterized (9) in terms of δ and α (the dependence on α is through D_{WSN}), we can determine the minimum number of sensors required for this WSN by finding the value of M such this inequality in (9) is satisfied. Part of this result is intuitive. It says that our minimum bound for this WSN depends on the relative sizes of the event area and sensors' sensing area. However, the dependence on δ is not intuitive, and it cannot be accounted for in the area coverage approach used in prior WSN literature. Fig. 17 illustrates the importance of δ by plotting the our bound for three different values of δ . For example, if the required probability of correct decision at the sink is 0.9, then $\delta = 1/2$ has bound 131 sensors, $\delta = 1/3$ has bound 106 sensors, and $\delta = 1/5$ has bound 64 sensors.

For comparison with the WSN Coverage problem from prior WSN research, we repeat a result from [25]. Let $P_{uncovered}$ equal the percent of A not covered by at least

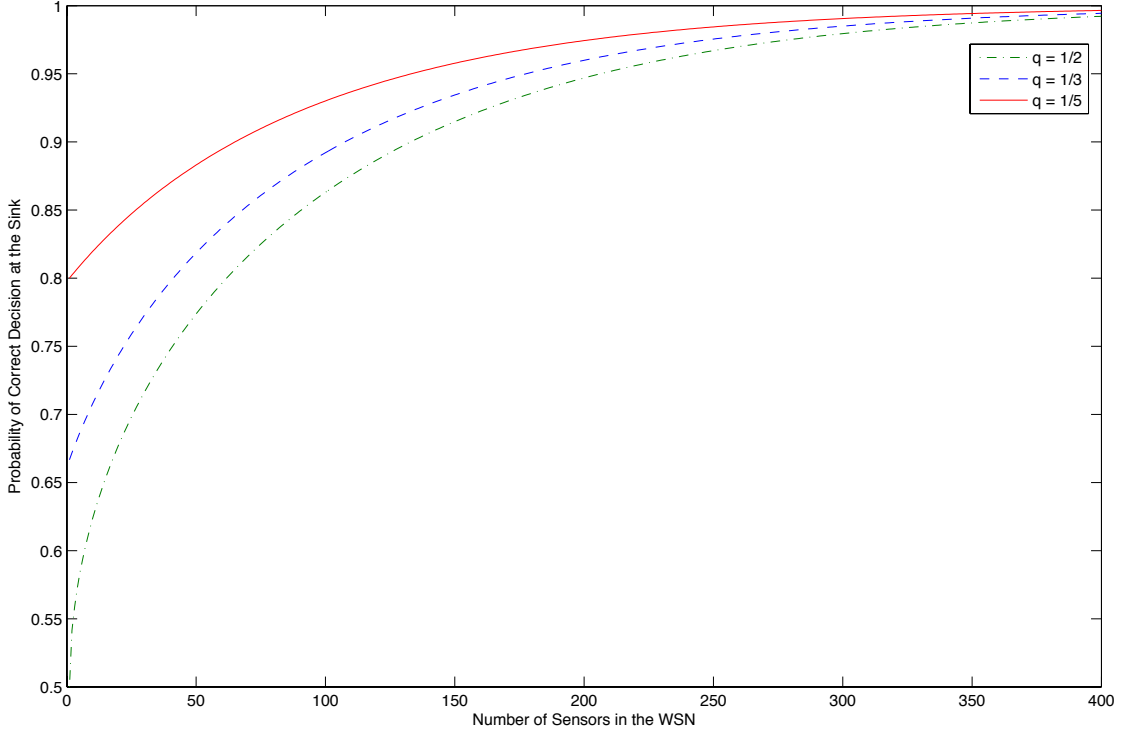


Figure 17: Information theoretic bound on the minimum number of sensors in the WSN for three values of δ and $\alpha = 1/100$.

one sensor. Then $P_{uncovered} = (1 - \alpha)^M$. The amount of uncovered area corresponds to the amount of distortion over a single time interval. Prior WSN research suggests that to design a WSN, the network operator should decide what percentage of time the event needs to be detected (which is equal to $1 - D$) and then design a WSN so that $P_{uncovered}$ is less than this percentage. In other words, the designer should make M such that

$$D \geq (1 - \alpha)^M. \quad (10)$$

Note how this compares to our bound in (9). In Fig. 18, we plot our results and the WSN Coverage results in terms of the number of sensors versus probability of the sink making a correct decision (i.e. $1 - D$). As shown in the figure, the information-theoretic approach predicts a much smaller bound on the number of sensors to achieve a given level of performance. For example, if the WSN must result in a correct decision 60% of the time, the coverage approach indicates the WSN must have 92 sensors while

the information theoretic approach indicates a bound of 6 sensors. Furthermore, if the WSN must result in a correct decision 95% of the time, the coverage approach indicates the WSN must have 231 sensors while information theory indicates a bound of 132 sensors.

To confirm the performance predicted by the information theoretic bound, Fig. 18 also plots simulation results for the WSN model in this example. The simulations generated 100 random WSNs for each value of M and simulated each WSN for 10,000 time intervals. Although the information theoretic bound only applies to the average WSN performance for each value of M , the plot also shows the maximum and minimum performance for each value of M . As seen in the figure, the simulations confirm the expected results: average simulated WSN performance never exceeds that predicted by the information theoretic bound (as long as edge effects are accounted for). Interestingly, the maximum simulated performance is often close to our bound. This further suggests that our bound could be an appropriate WSN design metric. The figure also shows that the minimum simulated performance is greater than $P_{covered}$ in most cases. This indicates that the area coverage approach from prior work will often lead to an inefficient WSN. The efficiency gained by using the information theoretic approach instead of the coverage approach is similar to the reduction in information rate achieved when performing source coding with a fidelity criterion instead of quantization.

Note that the primary factor in our bound outperforming traditional approaches is that our bound takes into account characteristics of the random process being estimated (in this case the characteristic is δ).

As a final comment, it is clear that the bound from the Negative WSN Coding Theorem is not tight. Therefore, we cannot use that bound to prove a positive coding theorem counterpart to Theorem 1. In Section 4.1.3 we try to modify the bound to make it tight.

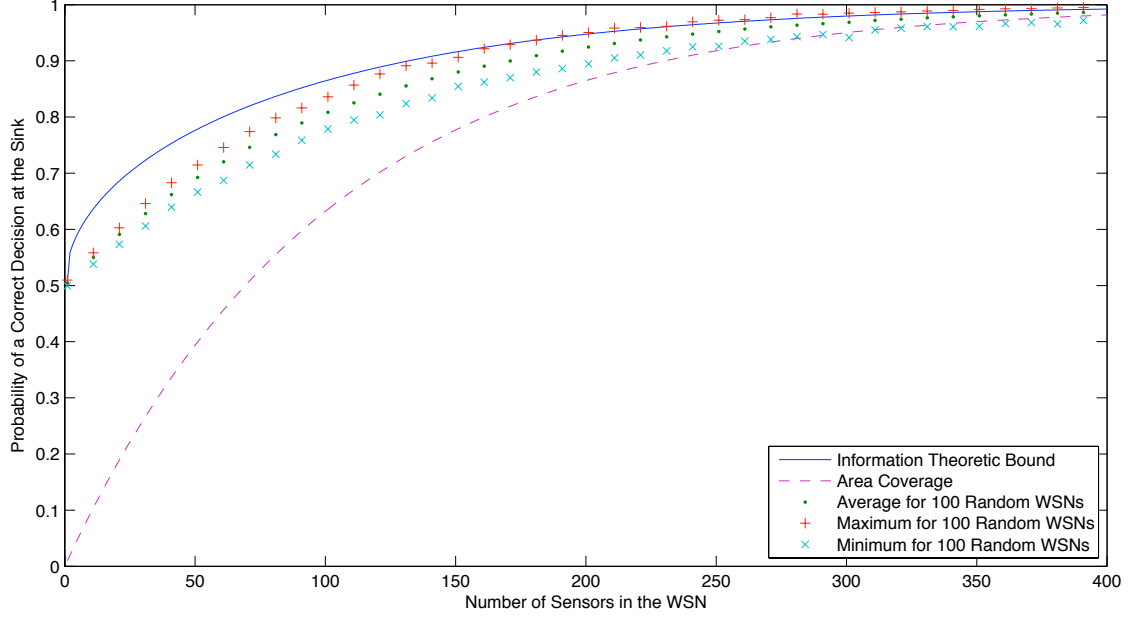


Figure 18: Performance comparison between the information theoretic bound developed in this work and the coverage approach implied by prior WSN research for $\delta = 1/2$ and $\alpha = 1/100$. Simulations used 10,000 time intervals and 100 random WSNs for each value of M .

4.1.2 WSN Extension to Fundamental Problem - Perspective 3

In order to answer the question regarding what types of events a previously-deployed WSN can accurately sense, we then take Perspective 3 on the WSN extension to the fundamental problem in information theory. This problem setup is shown in Figure 19.

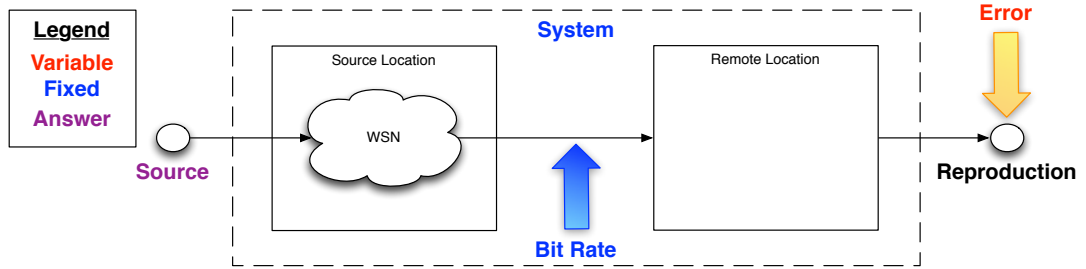


Figure 19: Perspective 3 of the WSN extension to the fundamental problem in information theory

As mentioned in Section 3.1, source coding theorems from Perspective 3 of the

fundamental problem in information theory tell us that a system cannot achieve the required distortion D if the system's bit rate is less than $R(D)$.

As before, even if we don't assume any structure on the WSN, we know that the useful bit rate of useful information being sent to the sink is limited by I_{WSN} , the information that the WSN collects about the source. This problem setup is shown in Figure 20.

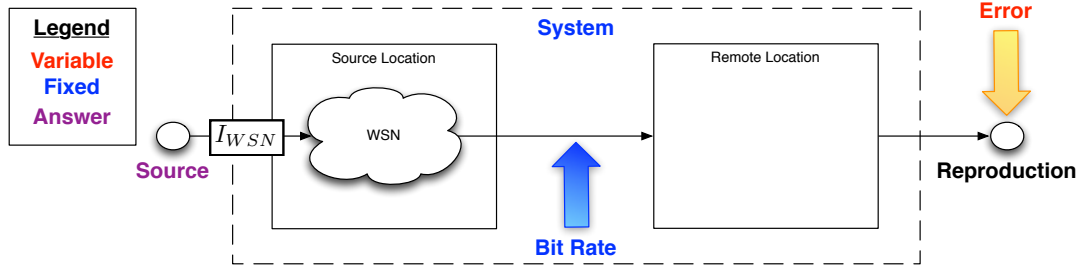


Figure 20: Adapting perspective 3 of the WSN extension to the fundamental problem in information theory

If the number of sensors in the WSN is fixed at M , we can define a *collection capacity* for the collection channel given by

$$C_{WSN} = \sup_X \bar{I}(X; S).$$

This collection capacity describes the maximum amount of information the collection channel can measure about an event. We can now use this fact to make C_{WSN} a top-level design metric to determine the classes of random processes the WSN can estimate. This is proved in the following Negative Collection Channel Coding Theorem.

Theorem 2 (Negative Collection Channel Coding Theorem) *Assume a given WSN has a stationary collection channel, a fixed number of sensors M , and transport capacity great enough to send all sensors' measurements reliably to the sink. Then if $R(D) \geq C_{WSN}$, the sink cannot form an estimate with the required distortion.*

In other words, the WSN does not have enough collection capacity for the sink to accurately estimate the event.

Proof: As in the proof of the previous theorem, we can use the data processing inequality to write

$$\bar{I}(G; Y) \leq \bar{I}(X; S) \leq \sup_X \bar{I}(X; S) = C_{WSN} \leq R(D).$$

From the definition of the rate-distortion function, we know that the sink's estimate must have distortion greater than D . ■

As with the Negative WSN Coding Theorem, we would like to prove the positive coding theorem counterpart to the Negative Collection Channel Theorem. In other words, we would like a collection capacity theorem that indicates the WSN will be able to accurately detect an event if C_{WSN} is large enough. Unfortunately, immediate inspection tells us that the theorem above does not prove this. Consider for example an L_i and S_i such that all sensors are located at the origin and can only detect an event if it occurs within a circle of radius r around the sensor. For all realistic sensing models, this WSN will have $C_{WSN} = \epsilon > 0$. Next assume the event of interest is any random process with $R(D) < \epsilon$ and that with probability one it occurs outside the circular area around the sensors. In this case, the WSN will not accurately estimate the event even though $R(D) \leq C_{WSN}$.

4.1.3 Positive Coding Theorems

As mentioned in Section 1.4, we would ideally like to prove we would like to prove both a positive and negative theorem for the collection channel as well as a positive and negative theorem for WSN coding. In this section, we discuss why it is not possible to do this in a meaningful way.

We have noted that part of the power of coding theorems from information theory is that they are generally applicable. That is, given a basic problem setup, we can

determine the performance limits for large groups of systems. At the other end of this spectrum, we could have a theorem that is very narrow and applies to only one system. In the narrow case, the theorem is not useful because, given a specific system, we can always calculate the performance limit and we do not need the theorem to provide the limit.

This fact presents a difficulty when considering the WSN extension to the fundamental problem. If possible, we would like to impose additional structure on the problem in order to make the bounds in Sections 4.1.1 and 4.1.2 tight. However, we do not want to restrict the problem so much that the class of systems for which the results apply is not useful. Unfortunately, there is no obvious way to restrict the WSN structure without eliminating options that would be useful for a WSN designer. For example, placing restrictions on \bar{f}_S would potentially eliminate sensor technologies that a designer could find useful. As a result, we do not attempt to further restrict the WSN structure in order to prove the positive counterparts of these theorems.

4.1.4 Summary of Top-Level Design Metrics

In this section we did two things. First we found a lower bound on the number of sensors the WSN must contain in order to accurately estimate ground truth. This bound is general and makes no assumptions on the WSN structure. Second, we found a lower bound on the collection capacity a WSN must have in order to accurately estimate certain classes of ground truth random processes. Again this bound is general, and makes no assumptions on the WSN structure. Unfortunately, in both cases, the bounds were not tight. As a result, they cannot have positive coding theorem counterparts. Furthermore, we showed that tight bounds (and associated positive coding theorems) do not appear to exist for general classes of WSNs.

4.2 WSN Sensing Efficiency Measure

As described in Section 3.1, the information theory results from Perspective 1 and Perspective 2 of the fundamental problem have indicated the best possible performance of a communication system. Alternatively, we can interpret these results as specifying the best possible performance of source or channel codes in a certain class. Results from Perspective 1 can be considered to apply to the class of codes with rate less than a specified value, and results from Perspective 2 can be considered to apply to the collection of codes with distortion less than a specified value.

For example, consider the class

$$\mathcal{C}_R = \{\text{Codes with distortion less than or equal to } D\}$$

in Perspective 2 of the fundamental problem. The rate-distortion function says that the minimum bit rate for any code in this class is $R(D)$.

When viewing the problem this way, coding theorems are often used to evaluate “how good” a specific code is relative to other codes in its class. This is why it is desirable for the coding theorems from information theory to hold for the most general classes possible.

Following the development in [24], we will use this approach to evaluate the performance of a WSN.

4.2.1 Efficiency Measure Defined

As discussed in Section 1.2, a given WSN with a specified number of sensors has an associated information rate with ground truth, denoted I_{WSN} , and an optimal average distortion achievable at the sink, denoted D_{WSN} . We can then consider the WSN as belonging to two classes of codes on the ground truth random process

$$\mathcal{C}_R = \{\text{Codes with rate less than or equal to } I_{WSN}\} \text{ and}$$

$$\mathcal{C}_D = \{\text{Codes with distortion greater than or equal to } D_{WSN}\}.$$

Since our model requires source coding theorems to hold for ground truth, Perspective 1 tells us that the minimum distortion of any code $C \in \mathcal{C}_R$ is given by the distortion rate function, $D(I_{WSN})$. Source coding theorems applied in Perspective 2 also tell us that the minimum rate of any code $C \in \mathcal{C}_D$ is given by the rate distortion function from (1). Therefore, a WSN's distortion can be no less than $D(I_{WSN})$ and a WSN's mutual information rate with ground truth can be no less than $R(D_{WSN})$.

Since $D(I_{WSN})$ and $R(D_{WSN})$ provide theoretical limits for WSN performance, we compare the actual performance of the WSN with this theoretical limit. We express the comparison as an efficiency measure written in terms of the WSN's percent decrease in performance relative to the theoretical limit. We can define the distortion efficiency as the excess distortion above the minimum, which is written

$$\eta_D = \frac{D_{WSN} - D(I_{WSN})}{D(I_{WSN})}. \quad (11)$$

We can also define the bit rate efficiency as the excess bit rate above the minimum, which is written

$$\eta_R = \frac{I_{WSN} - R(D_{WSN})}{R(D_{WSN})}. \quad (12)$$

Considering our focus on a WSN's sensing functionality, we can interpret these efficiency measures as follows. The WSN collects information about G through its sensors' measurements, thereby reducing the uncertainty in G . We can express the amount of information the WSN collects as I_{WSN} . The sink uses this collected information to develop a minimum-distortion estimate with accuracy D_{WSN} . The smallest amount of information for ANY system trying to estimate G with accuracy D_{WSN} is given by $R(D_{WSN})$. The quantity $I_{WSN} - R(D_{WSN})$ then describes the amount of information over the minimum possible that the WSN collects about G . Since it takes energy to collect information, having $I_{WSN} > R(D_{WSN})$ means there is an energy inefficiency within the WSN. To maximize the WSN's lifetime, we would like to minimize this inefficiency as much as possible.

4.2.2 Efficiency Analysis

Using this definition of efficiency, we analyze two simple examples of a WSN to understand the fundamental efficiency aspects of the WSN scenario. Our goals will be to determine if there are any causes of efficiency that are inherent to WSNs, and to also determine if there are characteristics that make one WSN more efficient than another.

We assume both example WSNs have M sensors deployed randomly and uniformly throughout a given geographic area to detect whether or not an event occurs. The first example is the same example used in Section 4.1. Even though it quite simple, it is commonly used in the WSN literature [25], and we can think of it as the best case scenario for a WSN since any inefficiencies in this WSN should be present in all other WSNs.

For this example, we already showed that

$$\begin{aligned} D_{WSN} &= \delta (1 - \alpha)^M, \\ I_{WSN} &= H_b(\delta) - (1 - \delta + D_{WSN}) H_b\left(\frac{1}{1 + \frac{D_{WSN}}{1 - \delta}}\right) \text{ and,} \\ R(D) &= H_b(\delta) - H_b(D). \end{aligned}$$

Therefore, it is straightforward to calculate that

$$\eta_R = \frac{H_b(D_{WSN}) - (1 - \delta + D_{WSN}) H_b\left(\frac{1}{1 + \frac{D_{WSN}}{1 - \delta}}\right)}{H_b(\delta) - H_b(D_{WSN})}. \quad (13)$$

We can see that η_R is completely specified by δ , α , and M . Since $D(R)$ is the inverse of $R(D)$, η_D is also completely specified by these parameters. Therefore, we might expect that each of these parameters has an effect on the WSN's η_R and η_D . Figures 21 and 22 illustrate the effects of δ , α , and M .

Because the plot of η_D is more intuitive than the plot of η_R , we use (13) to calculate and plot η_D instead of plotting (13). Note also that we use D_{WSN} as a proxy for M

in the figures. We do this since our perspective is of a WSN designer, and it is more natural for a WSN design to specify the required distortion of the WSN instead of specifying the required number of sensors. This does not affect the results since δ and α are fixed at a given point in the figures, and, as a result, D_{WSN} is completely specified for a given M .

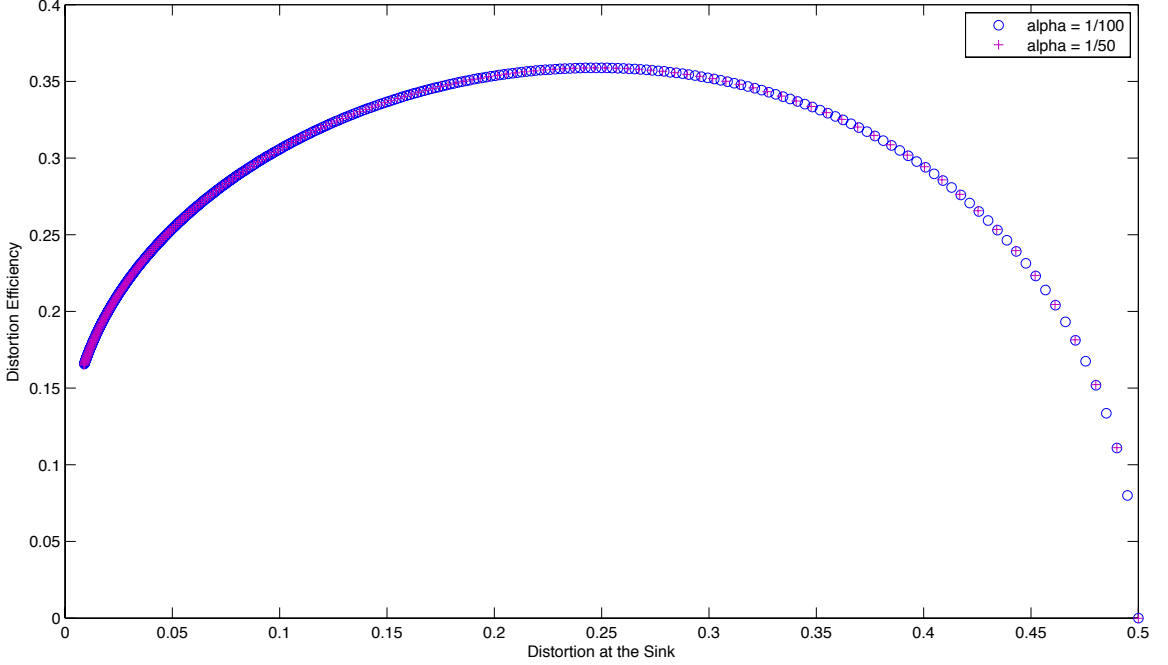


Figure 21: Distortion efficiency of a WSN for $\delta = 1/2$ and two values of α .

These figures indicate three things. First, the efficiency is extremely dependent on M (through the parameter D_{WSN}). We can see that in all cases, η_D decreases as D_{WSN} approaches its minimum value of 0 and as D_{WSN} approaches its maximum value of δ . Therefore, from a sensing functionality perspective, it makes more sense to use a WSN in applications that require large or small distortion since the WSN will have the best efficiency in those cases.

Second, the results show that η_D depends on the value of δ , and the WSN becomes less efficient as δ approaches $1/2$. Since these characteristics are outside the control of a WSN designer, their contribution to η_D can be thought of as a fixed amount of distortion that arises in the WSN application.

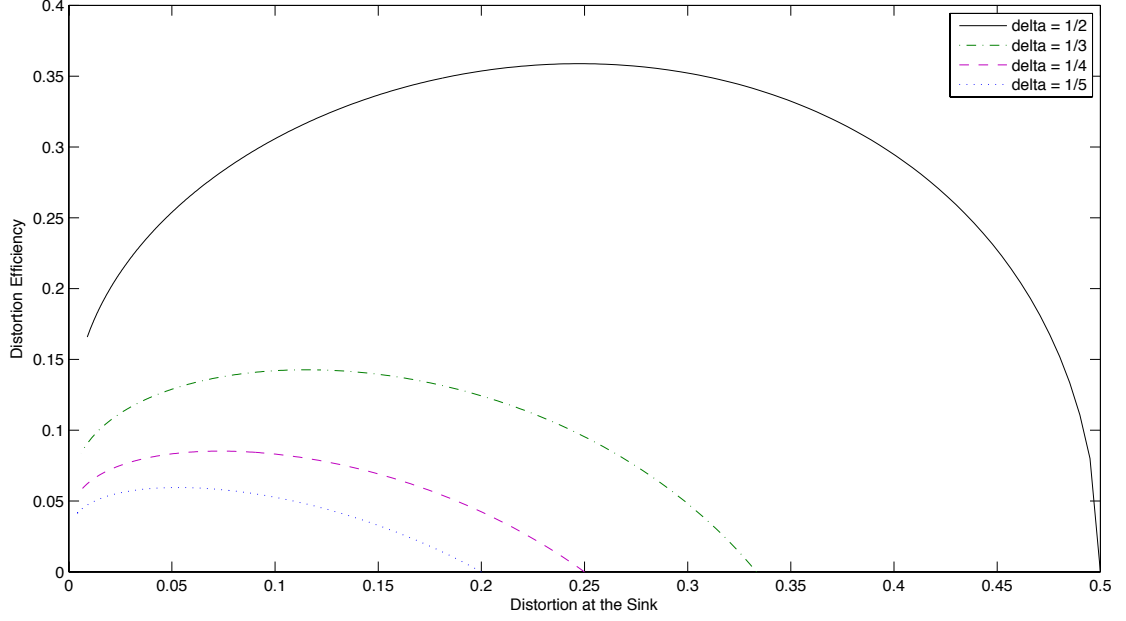


Figure 22: Distortion efficiency of a WSN for four values of δ and $\alpha = 1/100$.

Finally, we can see that the η_D is identical for both values of α , which indicates the WSN's efficiency is independent of the size of the area over which an individual sensor can take measurements. However, we can show that $D_{WSN} = \delta(1 - \alpha)^M = \delta V_A$, where V_A is the percentage of the geographic region that is not measured by any sensor. As a result, η_D also depends on V_A . Although this dependence may seem intuitive, the relationship between V_A and η_D is not. This is because greater V_A does not always result in less efficiency. In fact, the Figures 21 and 22 show that most values of η_D correspond to both a large and small value of V_A .

It might seem that the method of random sensor placement can cause some of this efficiency. However, it can be shown that, even if we place the sensors optimally so that they do not overlap and perfectly cover the geographic area, I_{WSN} is still given by (6). We can also see that we still have $D_{WSN} = \delta(1 - m\alpha) = \delta V_A$. This clearly shows that η_D still depends on V_A .

To further examine the causes of inefficiency, we consider a second example WSN. As before, a sensor can detect an event if the event occurs in a circular radius around

the sensor. However, we let the sensor make an error in its measurement with probability ϵ .

With this model, it can be shown that

$$I_{WSN} = H_b(\delta) - \sum H_b(q_s) p_s(s) \quad (14)$$

where $q_s = Pr(X = 0 | S = s)$. q_s is a complicated function of δ , α , ϵ , m , and s . The details of this calculation are omitted for brevity.

Using an approach similar to the previous WSN, the optimal estimation function can be shown to be

$$f_Y(s) = \operatorname{argmax}_x p_{S|X}(s | X = x). \quad (15)$$

Figure 23 shows η_D as a function of D_{WSN} for four values of ϵ . Note that $\epsilon = 0.0$ corresponds to the noiseless sensor case considered above. By looking at the Figure, we can see that, for many values of D_{WSN} , η_D is lower for the WSN with noisy sensors. As a result, we can infer that, for certain distortion levels and certain WSNs, the sensing functionality of a WSN with noisy sensors is more efficient than a WSN with noiseless sensors.

4.2.3 Summary of Efficiency Analysis

This analysis has showed two things. First, the parameters M , δ , and V_A introduce an inherent amount of inefficiency in the WSN's sensing functionality. Second, in certain scenarios noisy sensors can make a WSN's sensing process more efficient.

We conclude our analysis by noting that the efficiency of the sensing functionality of a WSN is a complicated function of the parameters in the WSN - especially the amount of uncovered area and the error in sensor measurements. Therefore, this efficiency should be carefully evaluated before choosing a technology to implement in the WSN. We also note that, although the results above suggest ways to design a WSN so that its sensing functionality is more efficient, the overall efficiency of a WSN is a combination of many factors in addition to the sensing functionality. Therefore,

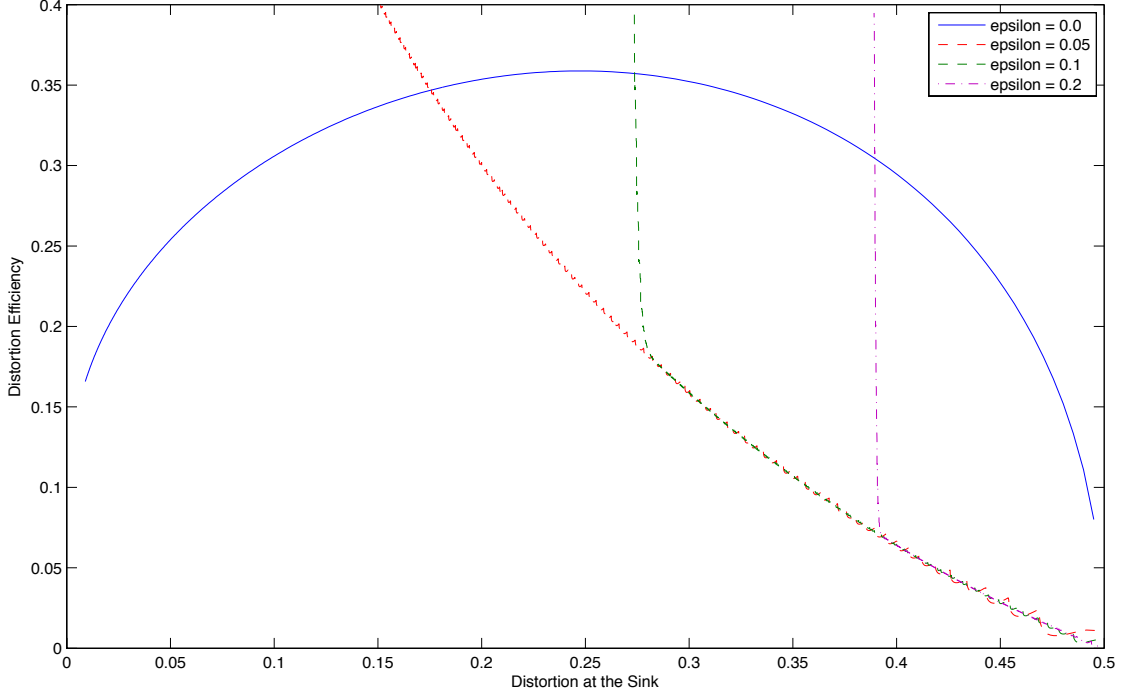


Figure 23: Distortion efficiency of a WSN for $\delta = 1/2$, $\alpha = 1/100$, and four values of ϵ .

the efficiency metric in this paper is only one component in designing an optimal, real-world WSN.

4.3 *New Sensor Design*

Sections 4.1 and 4.2 both showed that current WSNs do not currently perform as well as traditional coding from information theory. This leads to a natural question: “Can traditional information theory techniques/approaches give any hints about how to design better WSNs or sensors?”

We first note that the WSN approach to estimating ground truth is developing sensors that can accurately measure features of the event and send the measurements to the sink using the smallest transmission rate possible. In terms of sensing functionality, most efforts focus on sensing new features or sensing existing features more accurately. Given the fact that the real goal is a specified distortion between the sink’s estimate and ground truth, this is perhaps the wrong focus. As evidence of

this point, the efficiency measures in Section 4.2 indicate that we are collecting information about the event that is above and beyond the minimum required. In fact, this is why our example showed that noisy sensors are sometimes more efficient than noiseless sensors.

For comparison, we describe the coding approach used in source coding theorems from information theory. In this approach, the focus is not solely on the input signal. The focus is also on the minimum set of outputs needed to reproduce the source with sufficient accuracy. An example of this is illustrated in Figure 24 and described below.

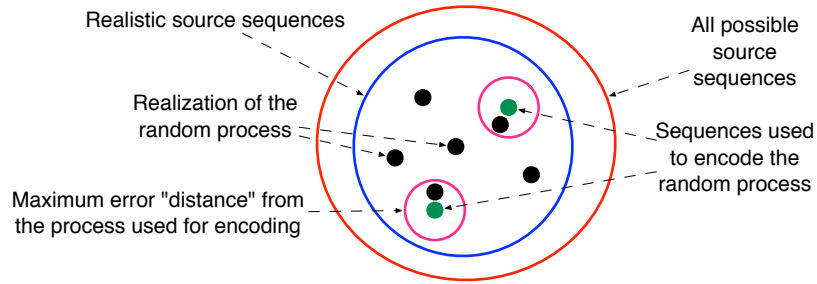


Figure 24: Illustration of Traditional Source Coding

Because of the technical conditions we require for G , we know that it satisfies the Asymptotic Equipartition Property (AEP) [18]. The AEP tells us that the set of all infinite-length sequences of G , given by $g_i = \{g_0, g_1, g_2, \dots\}$ can be separated into two sets: typical sequences and atypical sequences. If we look at any realization of the random process, with probability one the realization is a member of the typical sequences. Now, since there is zero probability that an atypical sequence can occur, our system only needs to worry about coding a typical sequence. To perform the encoding, our system will choose a subset, \mathcal{C} of the typical sequences and map any input it receives into the member of \mathcal{C} that is closest to the input. “Closeness” is measured relative to the fidelity criterion specified in the problem. Source coding

theorems tell us how big the set \mathcal{C} needs to be in order to achieve a required average distortion.

If we apply this approach to a WSN, WSN design would occur as follows. First we would consider the ground truth phenomena that must be reproduced and the fidelity criterion used to measure the reproduction's accuracy. Next we would determine the set \mathcal{C} that gives the optimal system performance. Then, the focus would be on developing the WSN so that it could efficiently determine which member of \mathcal{C} was closest to a given event sequence. This is the key difference from the current approach since the current approach focuses on estimating the event sequence at each individual time instant. This approach can be generalized so that various source coding mechanisms are considered and the WSN is designed to mimic the functionality of the mechanism that is most appropriate for the ground truth process under consideration.

Ultimately, this approach to WSN design could be more efficient since it focuses all of the design efforts on the primary objective of the WSN - reproducing a random process at a remote location.

CHAPTER V

SUMMARY

The goal of this thesis was to improve the design of WSNs using tools and techniques from information theory. With this approach, we used source coding theorems from information theory to develop three results that aid in WSN design. First, we proposed a new top-level design metric for WSNs. This metric indicates the minimum number of sensors needed for the WSN to function properly and also indicates the classes of random process that cannot be accurately estimated by a previously-deployed WSN. Our second result was an efficiency measure for the sensing process in a WSN. This measure helped identify fundamental causes of inefficiency in a WSN and was used to find conditions under which certain WSNs were more efficient. Finally, we used techniques from source coding schemes to suggest new designs for WSNs and the sensors they contain.

APPENDIX A

TECHNICAL CONDITIONS FOR THE WSN MODEL

In this chapter, we describe technical conditions on \bar{f}_G and \bar{f}_S that must hold for our WSN model. These conditions must hold so that source coding theorems from information theory apply.

Before describing the technical conditions, we explain some notation. We denote random variables with capital letters and their realization with lower case letters. We will let $P(\cdot)$ denote the probability distribution for a random variable, vector, or process when the subscript of P would just be a capitalized version of its arguments (i.e. $P(x) = P_X(x)$). Unless noted otherwise, given a random variable X , we will write the induced probability space by $(\Omega_X, \mathcal{B}(\Omega_X), P_X)$.

We use subscripts to denote a specific sample of a random process at either a time or sensor index. We use superscripts to indicate both a vector of samples of a random process as well as a vector of samples from different sensors. For example, t samples of the random process $\{G_t\}$ is written $G^t = \{G_0, \dots, G_{t-1}\}$ and t samples of the random process $\{S_{i,t}\}$ from M different sensors is written

$$\begin{aligned} S^{(M),t} &= \{S_{0,0}, \dots, S_{0,t-1}, \dots, S_{M-1,0}, \dots, S_{M-1,t-1}\} \\ &= \{S_0^{t-1}, \dots, S_{M-1}^{t-1}\} = \{S_0^{(M-1)}, \dots, S_{t-1}^{(M-1)}\}. \end{aligned}$$

We use the notation $S^{(M),t}$ when we want to emphasize explicitly the number of sensors in the WSN or the time index. However if we simply want to refer to all the sensors in the WSN without emphasizing the number of sensors, the time index, or either one, we use the notation S^t , $S^{(M)}$, or S . When the meaning is clear from the context, we will also use capital letters to refer to the appropriate random processes.

We let T^K denote the operator that shifts a random process K -units in time. In

other words, given a random process $X = (x_0, x_1, \dots)$, T^K is a function $T^K : \Omega_X^\mathcal{T} \rightarrow \Omega_X^\mathcal{T}$ where K is an integer and $T^K(x_1, x_2, \dots) = (x_{K+1}, x_{K+2}, \dots)$.

We require that the G induced by \bar{f}_G be stationary or asymptotically mean stationary (AMS) depending on the fidelity criterion used for the estimate at the sink. For practical purposes, this requirement also means that X must be stationary or AMS as well.

Note that any realistic \bar{f}_G will produce K samples of the ground truth sequence from N samples of the event sequence according to some deterministic function (e.g., the indicator function, identity function, average, maximum, minimum). We then can write G mathematically as $g_{jK}^K = f_G(T^{jN}x)$ where f_G is a deterministic function, and j , K , and N are integers. In this case, $\bar{f}_G = \{f_G(T^{jN}x); j \in \mathcal{T}\}$. As shown in [18] such a setup results in an AMS G if X is AMS. Therefore any realistic \bar{f}_G satisfies our requirement.

Similarly, we restrict \bar{f}_S to those functions that produce a stationary or AMS S when X is stationary or AMS, respectively. For example, consider the following generalization of virtually all realistic sensing models. First, assume that a sensor's measurement is corrupted by a noise process, W_i , that is independent of X . In this case we have $\bar{f}_S : \Omega_X^\mathcal{T} \times \Omega_W^\mathcal{T} \rightarrow \Omega_S^\mathcal{T}$. For example the sensor measurements might be an additive noise channel where $s_{t,i} = f_S(x_t, w_i) = x_t + w_{t,i}$. Second, assume the sensors produce K measurements using N samples of the event and noise sequences. We can then write S mathematically as $s_{jK,i}^K = f_S(T^{jN}x, T^{jN}w_i)$ where f_S is a deterministic function, and j , K , and N are integers. Finally, assume W is an i.i.d. process (or more generally a B-process [18]). In this case, $\bar{f}_S = \{f_S(T^{jN}x, T^{jN}w); j \in \mathcal{T}\}$. With these assumptions, we immediately see that \bar{f}_S is actually a primitive channel, and, as a result, \bar{f}_S produces an AMS S if X is AMS [31]. Therefore virtually all realistic sensing models satisfy our requirement.

Note that even if the sensors' measurements are deterministic and do not depend

on a noise process, we can simply let $W_{i,t}$ be a random variable that takes on the value 0 with probability 1. Therefore we always model $S_{i,t}$ as a primitive channel.

Using this information, we provide a more general definition of the collection channel.

Definition: Following the notation in [18], the *collection channel*

$$[(\Omega_X, \Omega_W), \nu, \Omega_S]$$

with input alphabet (Ω_X, Ω_W) and output alphabet Ω_S is a family of regular conditional probability measures

$$\{\nu_{x,w|L,A} : x \in \Omega_X^T, w \in \Omega_W^T\}$$

on the measurable space

$$(\Omega_S^T, \mathcal{B}(\Omega_S^T)).$$

Two primary concerns when considering different types of channels are whether they are stationary and whether they are ergodic. Because we made assumptions on \bar{f}_S such that the collection channel is primitive, it is well known that the collection channel is stationary (AMS, ergodic) with respect to any stationary (AMS, ergodic) source (X, W) [31]. This proves the following lemma:

Lemma 2 (Stationarity and Ergodicity) *If a WSN is such that L and A do not vary with time, \bar{f}_S satisfies the requirements above, and W is a B-process then the WSN has a stationary and ergodic collection channel.*

APPENDIX B

PROOFS

In this chapter we prove the Negative WSN Coding Theorem, the Minimum Bound Lemma, and the Negative Collection Channel Coding Theorem.

Theorem 3 (Negative WSN Coding Theorem) *Consider a WSN coding scheme $S : \Omega_X \rightarrow \Omega_{S^{(M)}}$ with coding rate (in terms of number of sensor measurements per source symbol) M . If the code induces a mutual information rate between ground truth and sensor measurements less than the rate distortion function for ground truth and the sink's estimate, the sink's estimate will not have the required accuracy. In other words, $I_{WSN} = \bar{I}(G; S^{(M)}) \leq R(D)$ implies $\lim_{t \rightarrow \infty} t^{-1} E[\rho_t(g^t, y^t)] \geq D$.*

Proof: We first show that the information rate between ground truth and the sink's estimate is less than the information rate between ground truth and the sensors' measurements. Since $G_t \rightarrow S_t^{(M)} \rightarrow Y_t$ forms a Markov Chain, a straightforward application of the Data Processing Inequality gives for all t

$$I(G^t; Y^t) \leq I(G^t; S^{(M),t}). \quad (16)$$

It is well known [35, Th. 3.19] that this implies the inequality in the desired result

$$\bar{I}(G; Y) = \limsup_{t \rightarrow \infty} t^{-1} I(G^t; Y^t) \leq \limsup_{t \rightarrow \infty} t^{-1} I(G^t; S^{(M),t}) = \bar{I}(G; S^{(M)}). \quad (17)$$

We can then use (17) to write for any $q_t \in Q_{t,D}$

$$\begin{aligned} R(D) &= \limsup_{t \rightarrow \infty} \inf_{q_t \in Q_{t,D}} t^{-1} I(G^t; Y^t) \\ &\leq \limsup_{t \rightarrow \infty} t^{-1} I(G^t; Y^t) = \bar{I}(G; Y) \leq \bar{I}(G; S^{(M)}). \end{aligned} \quad (18)$$

Because the source coding theorem holds and $\bar{I}(G; S^{(M)}) \leq R(D)$, $q_t \notin Q_{t,D}$. By definition, this fact implies our desired result: $\lim_{t \rightarrow \infty} t^{-1} E[\rho_t(g^t, y^t)] \geq D$. ■

Lemma 3 *Let M_{min} be the largest M such that $\bar{I}(G; S^{(M)}) \leq R(D)$. For any $M < M_{min}$, the sink will not be able to estimate ground truth with the required average distortion. In other words, $M < M_{min}$ implies $\lim_{t \rightarrow \infty} t^{-1} E[\rho_t(g^t, y^t)] \geq D$.*

Proof: First we show that the information rate between ground truth and the sensors' measurements is non-decreasing as a function of the number of sensors. Assume the number of sensors in the WSN increases from l to m . WSNs with l sensors and m sensors have mutual information $I(G^t; S^{(l),t})$ and $I(G^t; S^{(m),t})$ respectively. We can use the chain rule for mutual information and the non-negativity property of mutual information to write

$$\begin{aligned} I(G^t; S^{(m),t}) - I(G^t; S^{(l),t}) &= \sum_{i=0}^{m-1} I(G^t; S_i^t | S^{(i-1),t}) - \sum_{i=0}^{l-1} I(G^t; S_i^t | S^{(i-1),t}) \\ &= \sum_{i=l}^{m-1} I(G^t; S_i^t | S^{(i-1),t}) \geq 0. \end{aligned}$$

Therefore if $l < m$ then

$$I(G^t; S^{(l),t}) \leq I(G^t; S^{(m),t}).$$

We can then use [35, Th. 3.19] again to write

$$\bar{I}(G; S^{(l)}) \leq \bar{I}(G; S^{(m)}). \quad (19)$$

From (19) and the definition of M_{min} we know that for any M such that $M < M_{min}$ we can write $\bar{I}(G; S^{(M)}) \leq \bar{I}(G; S^{(M_{min})}) \leq R(D)$. From Theorem 1 we have seen that if $\bar{I}(G; S^{(M)}) \leq R(D)$, the sink will not be able to estimate ground truth with the required average distortion. Combining these facts tells us that for any $M < M_{min}$, the sink will not be able to estimate ground truth with the required average distortion. ■

Theorem 4 (Negative Collection Channel Coding Theorem) *Assume a given WSN has a stationary collection channel, a fixed number of sensors M , and transport*

capacity great enough to send all sensors' measurements reliably to the sink. Then if $R(D) \geq C_{WSN}$, the sink cannot form an estimate with the required distortion. In other words, the WSN does not have enough collection capacity for the sink to accurately estimate the event.

Proof: As in the proof of the previous theorem, we can use the data processing inequality to write

$$\bar{I}(G; Y) \leq \bar{I}(X; S) \leq \sup_X \bar{I}(X; S) = C_{WSN} \leq R(D).$$

From the definition of the rate-distortion function, we know that the sink's estimate must have distortion greater than D . ■

APPENDIX C

CALCULATIONS FOR THE EXAMPLE WSN

In this chapter, we provide calculate D_{WSN} and I_{WSN} for the example WSN in Section 4.1. For convenience, we repeat the problem setup.

Let the event model, X , be an independently and identically distributed (i.i.d.) point source so that the distribution of X_t equals X for all t . Furthermore, assume that the location of X is a random variable that is uniformly spread over the event area, A .

Since the WSN's function is to detect an event, we can describe *ground truth* as whether or not the amplitude of the event at time t , has crossed a threshold. In other words, $\bar{f}_G : \Omega_X \rightarrow \{0, 1\}$ so that

$$G_t = \bar{f}_G(\omega_x) = \bar{f}_G(b, l_E) = \begin{cases} 1 & \text{if } b > C \\ 0 & \text{if } b \leq C \end{cases} \quad (20)$$

where $b \in B$ is the event's amplitude and $l_E \in A$ is the event's location.

This induces a distribution for ground truth with probabilities

$$P_{G_t}(G_t = 1) = P_X(b > C) = \delta$$

$$P_{G_t}(G_t = 0) = 1 - \delta.$$

Because $\{X_t\}$ is i.i.d. with respect to t , so is $\{G_t\}$. Therefore, we can describe ground truth as a binary i.i.d. random process G where $Pr(g_t = 1) = \delta$.

Let the sensing model, S_i , be as follows. If the event occurs within a circular region centered around a sensor and the event's amplitude is greater than a certain value, the sensor will perfectly detect that the event occurred. Let A_i denote the

circular region around sensor. In other words, $\bar{f}_S : \Omega_X \rightarrow \{0, 1\}$ so that

$$S_{i,t} = \bar{f}_S(\omega_x) = \bar{f}_S(b, l_E) = \begin{cases} 1 & \text{if } b > C, \text{ and } l_E \in A_i \\ 0 & \text{otherwise} \end{cases}. \quad (21)$$

We assume the placement model L is such that the sensors are placed randomly, uniformly, and independently throughout the event area. Therefore $\{L_i\}$ are i.i.d with respect to i . Furthermore, we assume that A_i is small relative to the event area, and let $\alpha = A_i/A$. With these assumptions we know that $\Pr(l_E \in A_i) = \alpha$.

By definition, the mutual information between ground truth and the sensors measurements is written

$$I(G; S^{(M)}) = H(G) - H(G | S^{(M)}) \quad (22)$$

where

$$H(G | S^{(M)}) = \sum_{S^{(M)}} P(S^{(M)}) H(G | S^{(M)} = s^{(M)}). \quad (23)$$

Since the sensors can perfectly detect an event that occurs in A_i , we could immediately simplify (23) to a form that can be calculated. However, we will first develop an intermediate result that will be useful when considering WSNs where the sensors' sensing regions are i.i.d random variables.

Since the $\{L_i\}$ are i.i.d with respect to i , the $\{A_i\}$, are also i.i.d with respect to i . This means $P(s)$ only depends on the number of non-zero sensor measurements, S_i . Let $P_S(\tilde{S} = \tilde{s}^i)$ equal the probability that i sensors have non-zero measurements. Then for example, if only one S_i is non-zero then we can write

$$P_{\tilde{S}}(S = \tilde{s}^1) = P_S(S_0 = 1, S_1 = 0, \dots, S_{M-1} = 0) = P_S(S_i = 1, S_j = 0 \text{ for } j \neq i).$$

Using this result in (23) allows us to write

$$\begin{aligned}
H(G|S^{(M)}) &= \sum_{S^{(M)}} P(S^{(M)}) H(G|S^{(M)} = s^{(M)}) \\
&= \sum_{S^{(M)}} P_S(S^{(M)}) \sum_X P_{G|S^{(M)}}(g|s^{(M)}) \log P_{G|S^{(M)}}(g|s^{(M)}) \\
&= \sum_{i=0}^M \binom{M}{i} P_S(\tilde{S} = \tilde{s}^i) \sum_X P_{G|\tilde{s}^i}(g|\tilde{s}^i) \log P_{G|\tilde{s}^i}(g|\tilde{s}^i) \\
&= \sum_{i=0}^M \binom{M}{i} P_S(\tilde{S} = \tilde{s}^i) H(G|\tilde{S} = \tilde{s}^i).
\end{aligned} \tag{24}$$

This equation is valid for all WSNs with i.i.d sensing regions (e.g. when the sensor locations are i.i.d random variables and the sensing regions are constant). As a result for any WSN with the objective of detecting an event and with i.i.d sensing regions we can write

$$\begin{aligned}
I(G; S^{(M)}) &= H(G) - H(G|S^{(M)}) \\
&= H(G) - \sum_{i=0}^M \binom{M}{i} P_S(\tilde{S} = \tilde{s}^i) H(G|\tilde{S} = \tilde{s}^i).
\end{aligned} \tag{25}$$

Notice that the model in (21) immediately implies that the conditional probability of an event given that any sensor detected an event is equal to one. Mathematically we write this as $P(G = 1|\tilde{S} = \tilde{s}^i) = 1$ when $i \neq 0$. As a result, for $i \neq 0$, we know that $H(G|\tilde{S} = \tilde{s}^i) = 0$. We can then simplify (24) as follows:

$$\begin{aligned}
H(G|S^{(M)}) &= \sum_{i=0}^M \binom{M}{i} P_S(\tilde{S} = \tilde{s}^i) H(G|\tilde{S} = \tilde{s}^i) \\
&= P_S(\tilde{S} = \tilde{s}^0) H(G|\tilde{S} = \tilde{s}^0).
\end{aligned} \tag{26}$$

Notice that this result only depends on the sensors' measurements being noiseless. It does not depend on the WSN having i.i.d sensing regions. Therefore for any WSN with the objective of detecting an event and containing sensors with noiseless measurements we can write

$$\begin{aligned}
I(G; S^{(M)}) &= H(G) - H(G|S^{(M)}) \\
&= H(G) - P_S(\tilde{S} = \tilde{s}^0) H(G|\tilde{S} = \tilde{s}^0).
\end{aligned} \tag{27}$$

Assume A is large relative to A_i . Then, since the probability of the event's location is uniformly distributed throughout A , the probability of the event happening in S_i 's sensing range is simply $\alpha = \|A_i\| / \|A\|$ where $\|\cdot\|$ denotes the area of a two dimensional set. Using this notation we can write

$$\begin{aligned} P_S(\tilde{S} = \tilde{s}^0) &= P(G = 0) P(\tilde{S} = \tilde{s}^0 | G = 0) + P(G = 1) P(\tilde{S} = \tilde{s}^0 | G = 1) \\ &= (1 - \delta) + \delta (1 - \alpha)^M, \end{aligned}$$

$$P(G = 0, \tilde{S} = \tilde{s}^0) = P(G = 0) P(\tilde{S} = \tilde{s}^0 | G = 0) = 1 - \delta, \text{ and}$$

$$P(G = 1, \tilde{S} = \tilde{s}^0) = P(G = 1) P(\tilde{S} = \tilde{s}^0 | G = 1) = \delta (1 - \alpha)^M.$$

We can then determine the other two probabilities required to calculate (27)

$$P(G = 0 | \tilde{S} = \tilde{s}^0) = \frac{P(G = 0, \tilde{S} = \tilde{s}^0)}{P_S(\tilde{S} = \tilde{s}^0)} = \frac{1 - \delta}{(1 - \delta) + \delta (1 - \alpha)^M} \quad (28)$$

and

$$P(G = 1 | \tilde{S} = \tilde{s}^0) = \frac{P(G = 1, \tilde{S} = \tilde{s}^0)}{P_S(\tilde{S} = \tilde{s}^0)} = \frac{\delta (1 - \alpha)^M}{(1 - \delta) + \delta (1 - \alpha)^M}. \quad (29)$$

Substituting these values into (27) gives

$$I_{WSN} = H_b(\delta) - (1 - \delta + D_{WSN}) H_b\left(\frac{1}{1 + \frac{D_{WSN}}{1 - \delta}}\right). \quad (30)$$

With an optimal estimation function, it is easy to see that

$$D_{WSN} = P(G = 1, \tilde{S} = \tilde{s}^0) = \delta (1 - \alpha)^M. \quad (31)$$

This concludes the results to be proved in this section.

REFERENCES

- [1] ADLAKHA, S. and SRIVASTAVA, M., “Critical density thresholds for coverage in wireless sensor networks,” in *Proc. IEEE Wireless Communications and Networking Conference (WCNC 2003)*, 2003.
- [2] AHMADI, H., “Probabilistic coverage and connectivity in wireless sensor networks,” Master’s thesis, Sharif University of Technology, 2005.
- [3] AHMED, N., KANHERE, S. S., and JHA, S., “Probabilistic coverage in wireless sensor networks,” in *Proc. of the IEEE Workshop on Wireless Local Networks (WLN, in conjunction with LCN 2005)*, Nov. 2005.
- [4] AMMARI, H. and DAS, S., “Critical density for coverage and connectivity in three-dimensional wireless sensor networks using continuum percolation,” *Parallel and Distributed Systems, IEEE Transactions on*, June 2009.
- [5] BERGER, T., ZHANG, Z., and VISWANATHAN, H., “The CEO problem [multiterminal source coding],” *IEEE Transactions on Information Theory*, May 1996.
- [6] BERGER, T., *Rate Distortion Theory*. Information and System Sciences Series, Englewood Cliffs, N.J.: Prentice-Hall, 1971.
- [7] BERGER, T., “Multiterminal source coding,” in *The Information Theory Approach to Communications* (LONGO, G., ed.), Springer-Verlag, 1978.
- [8] BHATTACHARYA, S., XING, G., LU, C., GRUIA-CATALIN, R., CHIPARA, O., and HARRIS, B., “Dynamic wake-up and topology maintenance protocols with spatiotemporal guarantees,” in *Proc. Fourth International Symposium on Information Processing in Sensor Networks (IPSN 2005)*, Apr. 2005.
- [9] BOUKERCHE, A., FEI, X., and ARAUJO, R. B., “An energy aware coverage-preserving scheme for wireless sensor networks,” in *Proc. of the Second ACM international workshop on Performance evaluation of wireless ad hoc, sensor, and ubiquitous networks (PE-WASUN ’05)*, (New York, NY, USA), ACM Press, 2005.
- [10] CARDEI, M. and WU, J., “Coverage in wireless sensor networks,” in *Handbook of Sensor Networks* (ILYAS, M. and MAHGOUB, I., eds.), CRC Press, 2004.
- [11] CHAMAM, A. and PIERRE, S., “On the planning of wireless sensor networks: Energy-efficient clustering under the joint routing and coverage constraint,” *IEEE Transactions on Mobile Computing*, Aug. 2009.

- [12] CHAMBERLAND, J. F. and VEERAVALLI, V. V., "Decentralized detection in sensor networks," *IEEE Transactions on Signal Processing*, Feb. 2003.
- [13] COVER, T. M. and THOMAS, J. A., *Elements of Information Theory*. Hoboken, NJ: Wiley, 2006.
- [14] DAS, A. and KEMPE, D., "Sensor selection for minimizing worst-case prediction error," in *IPSN '08: Proceedings of the 7th international conference on Information processing in sensor networks*, (Washington, DC, USA), IEEE Computer Society, 2008.
- [15] GAO, Q., ZHANG, J., LARISH, B., and SHEN, S., "Coalition-aided data transmissions in wireless sensor networks," in *Communications, 2006. ICC '06. IEEE International Conference on*, Jun. 2006.
- [16] GASTPAR, M. and VETTERLI, M., "Power, spatio-temporal bandwidth, and distortion in large sensor networks," *IEEE Journal on Selected Areas in Communications*, Apr. 2005.
- [17] GRAY, R.M. NEUHOFF, D., "Quantization," *Information Theory, IEEE Transactions on*, Oct. 1998.
- [18] GRAY, R. M., *Entropy and Information Theory*. New York, NY: Springer, 2011.
- [19] HUANG, C.-F. and TSENG, Y.-C., "A survey of solutions to the coverage problem in wireless sensor networks," *Journal of Internet Technology*, 2005.
- [20] KIEFFER, J., "A survey of the theory of source coding with a fidelity criterion," *Information Theory, IEEE Transactions on*, Sep. 1993.
- [21] KUMAR, S., LAI, T. H., and BALOGH, J., "On k-coverage in a mostly sleeping sensor network," in *Proc. of the 10th annual international conference on Mobile computing and networking (MobiCom '04)*, (New York, NY, USA), ACM Press, 2004.
- [22] LARISH, B. and RILEY, G., "The collection channel in a wireless sensor network," in *Information Theory Workshop (ITW), 2010 IEEE*, 2010.
- [23] LARISH, B. and RILEY, G., "An information theoretic approach for determining the minimum number of sensors in a wireless sensor network," in *Information Theory Workshop (ITW), 2011 IEEE*, Oct. 2011.
- [24] LARISH, B. and RILEY, G., "Measuring the efficiency of the sensing process in a wireless sensor network," in *International Conference on Computing, Networking, and Communications (ICNC), 2013 IEEE*, Jan. 2013.
- [25] LIU, B. and TOWSLEY, D., "A study on the coverage of large-scale sensor networks," in *Proc. The 1st IEEE International Conference on Mobile Ad-hoc and Sensor Systems*, 2004.

- [26] LIU, X. and MOHAPATRA, P., “On the deployment of wireless sensor nodes,” in *Third International Workshop on Measurement, Modeling, and Performance Analysis of Wireless Sensor Networks*, Jul. 2005.
- [27] LUO, Z.-Q., “Universal decentralized estimation in a bandwidth constrained sensor network,” *IEEE Transactions on Information Theory*, June 2005.
- [28] MARCO, D., DUARTE-MELO, E., LIU, M., and NEUHOFF, D., “On the many-to-one transport capacity of a dense wireless sensor network and the compressibility of its data,” in *Information Processing in Sensor Networks*, Springer Berlin/Heidelberg, 2003.
- [29] MARCO, D. and NEUHOFF, D. L., “Entropy of quantized data at high sampling rates,” in *Information Theory, 2005. ISIT 2005. Proceedings. International Symposium on*, Sep. 2005.
- [30] MEGERIAN, S., KOUSHANFAR, F., QU, G., VELTRI, G., and POTKONJAK, M., “Exposure in wireless sensor networks: theory and practical solutions,” *Wireless Networks*, 2002.
- [31] NEUHOFF, D. and SHIELDS, P., “Channels with almost finite memory,” *Information Theory, IEEE Transactions on*, Jul. 1979.
- [32] PARK, S.-J., *Energy-aware topology control and data delivery in wireless sensor networks*. PhD thesis, Georgia Institute of Technology, Jul. 2004.
- [33] RACHLIN, Y., NEGI, R., and KHOSLA, P., “Sensing capacity for target detection,” in *Information Theory Workshop, 2004. IEEE*, Oct. 2004.
- [34] RACHLIN, Y., NEGI, R., and KHOSLA, P., “Sensing capacity for discrete sensor network applications,” in *Information Processing in Sensor Networks, 2005. IPSN 2005. Fourth International Symposium on*, Apr. 2005.
- [35] RUDIN, W., *Principles of Mathematical Analysis*. McGraw-Hill, 1976.
- [36] SHEN, F., SUN, M.-T., LIU, C., and SALAZAR, A., “Coverage-aware sleep scheduling for cluster-based sensor networks,” in *Wireless Communications and Networking Conference, 2009. WCNC 2009. IEEE*, (Budapest), Apr. 2009.
- [37] TSITSIKLIS, J. N., “Decentralized detection,” in *Advances in Signal Processing* (POOR, H. V. and THOMAS, J. B., eds.), JAI Preess, 1993.
- [38] UNSER, M., “Sampling-50 years after shannon,” *Proceedings of the IEEE*, Apr. 2000.
- [39] WANG, B., CHUA, K. C., SRINIVASAN, V., and WANG, W., “Information coverage in randomly deployed wireless sensor networks,” *Wireless Communications, IEEE Transactions on*, Aug. 2007.

- [40] XIAO, J.-J., CUI, S., LUO, Z.-Q., and GOLDSMITH, A., “Joint estimation in sensor networks under energy constraints,” in *Sensor and Ad Hoc Communications and Networks, 2004. IEEE SECON 2004. 2004 First Annual IEEE Communications Society Conference on*, Oct. 2004.
- [41] XU, K., *Device Deployment Strategies for Large-scale Wireless Sensor Networks*. PhD thesis, Queen’s University, 2008.